

ePub^{WU} Institutional Repository

Sylvia Frühwirth-Schnatter

Monitoring von ökologischen und biometrischen Prozessen mit statistischen Filtern

Working Paper

Original Citation:

Frühwirth-Schnatter, Sylvia (1991) Monitoring von ökologischen und biometrischen Prozessen mit statistischen Filtern. *Forschungsberichte / Institut für Statistik*, 18. Department of Statistics and Mathematics, WU Vienna University of Economics and Business, Vienna.

This version is available at: <http://epub.wu.ac.at/560/>

Available in ePub^{WU}: July 2006

ePub^{WU}, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.

Monitoring von ökologischen und biometrischen Prozessen mit statistischen Filtern



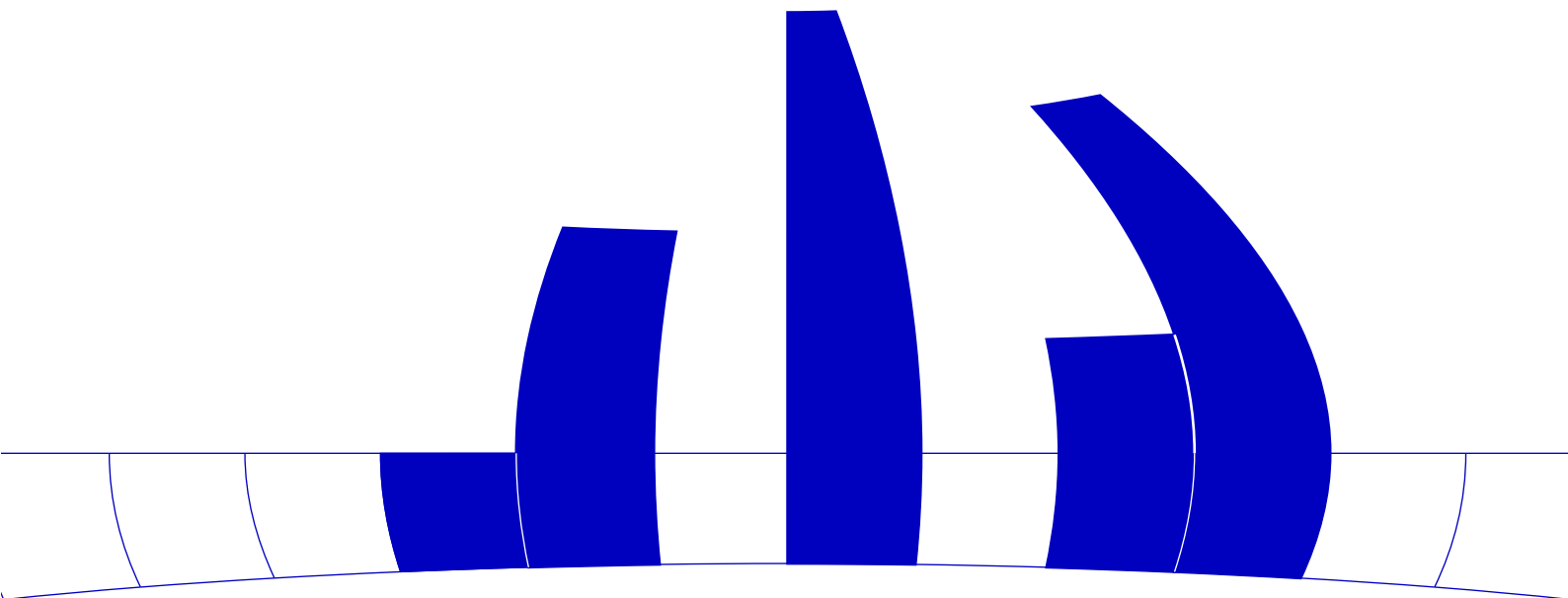
Sylvia Frühwirth-Schnatter

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 18
1991

<http://statmath.wu-wien.ac.at/>



Monitoring von ökologischen und biometrischen Prozessen mit statistischen Filtern

Sylvia Frühwirth-Schnatter
Institut für Statistik, Wirtschaftsuniversität Wien

Zusammenfassung

Diese Arbeit ist ein Überblick über die Ideen und Methoden der dynamischen stochastischen Modellierung von normalverteilten und nicht-normalverteilten Prozessen. Nach einer Einführung der allgemeinen Modellform werden Ausagemöglichkeiten wie Filtern, Glätten und Vorhersagen diskutiert und das Problem der Identifikation unbekannter Hyperparameter behandelt. Die allgemeinen Ausführungen werden an zwei Fallstudien, einer Zeitreihe des mittleren jährlichen Grundwasserspiegels und einer Zeitreihe von Tagesmittelwerten von SO_2 -Emissionen illustriert.

Schlüsselworte: Data-Augmentation, dynamische stochastische Modelle, dynamisches Trendmodell, Filtern, Gauß-Hermite-Integration, Glätten, Kalman-Filter, Monitoring, Multi-Prozeß-Filter, Steady-State-Modell, Trendanalyse, Vorhersagen.

1 Einleitung

Gegenstand dieser Arbeit bilden Prozesse, die durch regelmäßige Beobachtung einer meßbaren Größe y_t laufend erfaßt werden, um Aussagen über den Verlauf des Prozesses zu ermöglichen. Dieses Monitoring von Prozessen über Beobachtungsverläufe wird an zwei Zeitreihen aus dem Bereich der Ökologie illustriert.

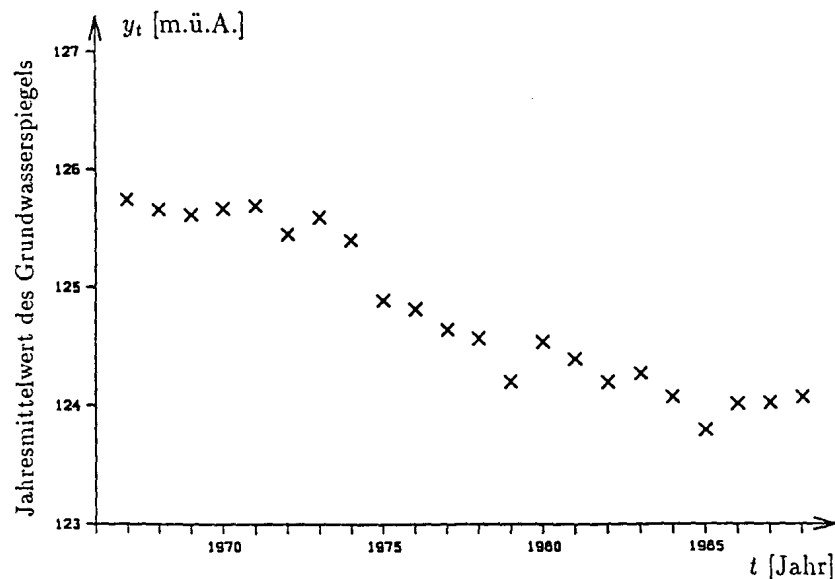
Datensatz 1 - Trendanalyse von Grundwasserdaten

Tabelle 1 enthält die Jahresmittelwerte des Grundwasserspiegels einer Meßstelle im Seewinkel im Burgenland (Österreich) von 1967 bis 1988. Diese Zeitreihe ist in Abbildung 1 graphisch dargestellt. Bei diesen Daten werden wir eine Aussage über die systematische Veränderung des Jahresmittelwertes des Grundwasserspiegels treffen.

Tabelle 1: Grundwasserspiegel y_t [m.ü.A.] einer Meßstelle im Seewinkel (Österreich)
(Jahresmittelwerte 1967 – 1988)

1967	1968	1969	1970	1971	1972	1973	1974	1975	1976
124.640	125.748	125.666	125.620	125.676	125.701	125.462	125.601	125.405	124.896
1977	1978	1979	1980	1981	1982	1983	1984	1985	1986
124.822	124.568	124.203	124.541	124.399	124.199	124.270	124.074	123.796	124.019
1987	1988								
124.028	124.070								

Abbildung 1: Graphische Darstellung von Datensatz 1
(siehe Tabelle 1)



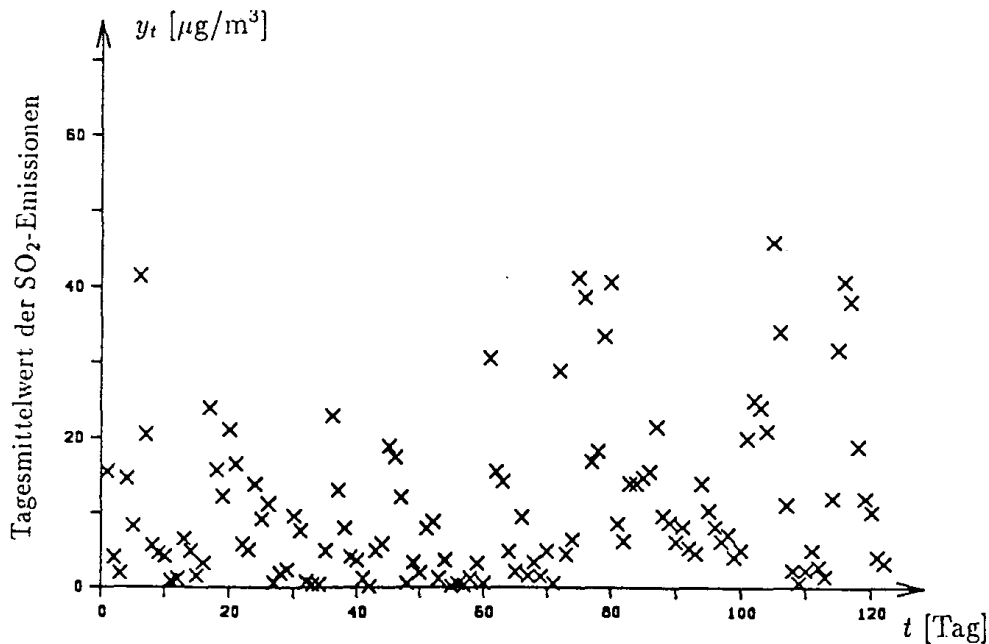
Datensatz 2 - Monitoring von SO₂-Emissionen

Tabelle 2 enthält die Tagesmittelwerte der SO₂-Emmissionen einer Meßstelle in Brotjachtriegel (BRD) vom 1.9.1976 bis zum 31.12.1976. Diese Zeitreihe ist in Abbildung 2 graphisch dargestellt. Bei diesem Datensatz werden wir für jeden Tag eine Aussage über die Wahrscheinlichkeit treffen, mit der am nächsten Tag ein bestimmter Schwellwert der SO₂-Belastung überschritten wird.

Tabelle 2: SO₂-Emissionen y_t [$\mu\text{g}/\text{m}^3$] einer Meßstelle in Brotjachtriegel (BRD)
(Tagesmittelwerte 1.IX.1976 – 31.XII.1976)

t (IX)	y_t	t (X)	y_t	t (XI)	y_t	t (XII)	y_t
1	15.5	31	7.7	62	15.5	92	5.3
2	4.0	32	0.8	63	14.2	93	4.6
3	1.9	33	0.4	64	4.8	94	14.0
4	14.6	34	0.4	65	2.1	95	10.4
5	8.3	35	4.9	66	9.5	96	8.2
6	41.4	36	22.8	67	1.6	97	6.3
7	20.5	37	13.0	68	3.4	98	7.2
8	5.7	38	8.0	69	1.5	99	4.1
9	4.6	39	4.1	70	4.9	100	5.1
10	4.2	40	3.6	71	0.6	101	20.0
11	0.9	41	1.2	72	28.9	102	25.0
12	1.2	42	0.1	73	4.5	103	24.1
13	6.6	43	4.9	74	6.5	104	21.0
14	4.9	44	5.8	75	41.2	105	45.9
15	1.6	45	18.8	76	38.6	106	34.1
16	3.3	46	17.4	77	16.9	107	11.2
17	23.9	47	12.0	78	18.3	108	2.3
18	15.7	48	0.5	79	33.4	109	0.7
19	12.2	49	3.3	80	40.6	110	2.3
20	21.0	50	1.9	81	8.6	111	5.0
21	16.5	51	7.9	82	6.3	112	2.8
22	5.8	52	8.8	83	14.0	113	1.5
23	5.0	53	1.2	84	14.0	114	12.1
24	13.8	54	3.6	85	14.8	115	31.8
25	9.2	55	0.2	86	15.6	116	40.8
26	11.2	56	0.5	87	21.5	117	38.2
27	0.7	57	0.3	88	9.7	118	19.0
28	1.9	58	1.2	89	8.8	119	12.1
29	2.4	59	3.2	90	6.2	120	10.2
30	9.6	60	0.6	91	8.3	121	4.1
		61	30.5			122	3.3

Abbildung 2: Graphische Darstellung von Datensatz 2



Die Modellklasse, die in dieser Arbeit zur Bewertung von Aussagen über Zeitreihen Anwendung findet, wurde unter der Bezeichnung „dynamic generalized linear model“ von [33] vorgeschlagen und ist im Detail in den Monographien von [13] und [32] behandelt. Wir werden für diese Modellklasse die Bezeichnung dynamische stochastische Modelle wählen.

Ziel der vorliegenden Arbeit ist es, Substanzwissenschaftlern wie Biometrikern oder Ökologen die Ideen und Methoden der dynamischen stochastischen Modellierung näher zu bringen. Dabei werden einerseits die bereits klassischen Ergebnisse für normalverteilte Prozesse zusammengefaßt und an den Grundwasserdaten illustriert. Andererseits wird die relativ spärlich untersuchte Problematik der dynamischen stochastischen Modellierung von nicht-normalverteilten Prozessen ausführlich diskutiert und an den SO₂-Daten illustriert.

Die Arbeit gliedert sich in 4 Abschnitte. In Abschnitt 2 wird die dynamische stochastische Modellierung von Zeitreihen an zwei einfachen Modellformen motiviert. In Abschnitt 3 werden die verschiedenen Analysemöglichkeiten wie Filtern, Glätten und Vorhersagen beschrieben, wobei wir bei nicht-normalverteilten Prozessen nicht den Vorschlägen von [33] folgen, sondern auf den Ideen in [9], [27] und [26] aufbauen. Abschnitt 4 behandelt das Problem unbekannter Hyperparameter. Mit einer in [11] vorgeschlagenen Methode wird die a-posteriori-Dichte der Parameter approximiert und die Analyse mittels eines Multi-Prozeß-Filters ([14]) durchgeführt. In Abschnitt 5 werden kurz Methoden der Modelldiagnose behandelt. Jeder Abschnitt enthält Fallstudien zu den beiden Datensätzen.

2 Dynamische stochastische Modellierung

2.1 Allgemeine Bemerkungen

Der erste Schritt der dynamischen stochastischen Modellierung besteht in der Annahme, daß die einzelnen Werte der beobachteten Zeitreihe Realisationen eines stochastischen Prozesses y_t sind. Zu jedem Zeitpunkt t wird die stochastische Variation der Werte, die dieser Prozeß annimmt, mit einer Wahrscheinlichkeitsverteilung einer bestimmten Verteilungsfamilie beschrieben. Die Parameter dieser Verteilung können einer Veränderung in der Zeit unterliegen, sodaß auch nicht-stationäre Prozesse direkt modellierbar sind.

Die Wahl der Familie hängt vom Charakter des beobachteten Prozesses ab. Bei den meisten Anwendungen wird angenommen, daß der Prozeß normalverteilt ist; in vielen Anwendungen ist diese Annahme auch tatsächlich gerechtfertigt. Wir werden im folgenden die Grundwasserdaten mit einer Normalverteilung modellieren (siehe Abschnitt 2.3.1).

Bei biometrischen oder ökologischen Zeitreihen stößt man mit der Normalverteilungsannahme mitunter an Grenzen, etwa wenn man Zeitreihen von Zählraten (z.B. Mortalitätsdaten) oder Stunden- oder Tagesmittelwerte von positiven metrischen Merkmalen, die nahe bei 0 liegen (z.B. Niederschläge oder Schadstoffemissionen), analysiert. Für solche Fälle wurde von [33] in Anlehnung an verallgemeinerte lineare Modelle ([17]) vorgeschlagen, mit allgemeineren Verteilungen zu arbeiten. Im folgenden modellieren wir den Prozeß der SO_2 -Emissionen mit einer Gamma-Verteilung (siehe Abschnitt 2.3.2).

Die Verteilung von y_t bei bekanntem Erwartungswert μ_t (in Zeichen $y_t|\mu_t$) wird im weiteren als Beobachtungsverteilung bezeichnet. [33] läßt als Beobachtungsverteilung nur Verteilungen der Exponentialfamilie zu. Für die in dieser Arbeit verwendeten Analysemethoden ist es möglich, jede Verteilung als Beobachtungsverteilung zu wählen, deren Erwartungswert $E(y_t|\mu_t)$ und Varianz $V(y_t|\mu_t)$ existieren und von folgender Gestalt sind:

$$E(y_t|\mu_t) = \mu_t, \quad V(y_t|\mu_t) = v_t(\mu_t) \cdot \phi. \quad (1)$$

$v_t(\cdot)$ ist eine positive Funktion von μ_t . ϕ ist ein positiver Parameter.

Der zweite Schritt der dynamischen stochastischen Modellierung besteht in der Beschreibung der Veränderung zwischen den Erwartungswerten μ_t und μ_{t-1} durch eine dynamische stochastische Gleichung. Ein Modell der Form

$$\mu_t = \mu_{t-1} + a \quad (2)$$

ist dynamisch und in folgendem Sinne deterministisch: sind μ_{t-1} und a bekannt, so ist der bedingte Erwartungswert von $(\mu_t|\mu_{t-1}, a)$ eine Größe, deren Wert mit Sicherheit vorhergesagt werden kann. Die Erweiterung bei der dynamischen stochastischen Modellierung besteht darin, die vorhersehbare Größe μ_t durch eine stochastische Größe zu ersetzen.

Für normalverteilte Prozesse schlug [14] eine Reihe solcher Modelle vor. Im folgenden Abschnitt beschreiben wir zwei dieser Modelle und verallgemeinern sie auf nicht-

normalverteilte Prozesse. Dieser Abschnitt enthält auch die allgemeine Modellform. Im Abschnitt 2.3 schlagen wir für die Datensätze 1 und 2 jeweils ein dynamisches stochastisches Modell vor.

2.2 Beispiele und die allgemeine Modellform

2.2.1 Das Steady-State-Modell

Das einfachste dynamische stochastische Modell beruht auf der Vorstellung, daß der Erwartungswert μ_t um den Erwartungswert μ_{t-1} nach einer Normalverteilung schwankt (Steady-State-Modell, [14]):

$$\mu_t = \mu_{t-1} + w_t, \quad w_t \sim N(0, Q). \quad (3)$$

Für $Q > 0$ ist diese Beziehung eine dynamische stochastische Gleichung. Bei bekanntem μ_{t-1} ist μ_t eine stochastische Größe, die mit Erwartungswert μ_{t-1} und Varianz Q normalverteilt ist. Je größer Q , desto stärker schwankt μ_t um μ_{t-1} .

Dieses Modell wurde von [14] für normalverteilte Prozesse vorgeschlagen und kann nicht direkt auf nicht-normalverteilte Prozesse angewendet werden, da der Erwartungswert μ_t im allgemeinen Fall nicht alle reellen Zahlen annimmt. Wird als Verteilung des Prozesses z.B. eine Gamma-Verteilung angenommen, so enthält der für μ_t zulässige Bereich E alle positiven reellen Zahlen. Gleichung (3) könnte zu negativen Werten führen. Je näher μ_{t-1} bei 0 liegt, desto unsymmetrischer müssen die Schwankungen von μ_t um μ_{t-1} sein. In Anlehnung an [33] beschreiben wir die Veränderung von μ_t auf einer transformierten Ebene:

$$g(\mu_t) = g(\mu_{t-1}) + w_t, \quad w_t \sim N(0, Q). \quad (4)$$

Die Transformation $g(\cdot)$ wird so gewählt, daß sie den für μ_t zulässigen Bereich E auf die reellen Zahlen abbildet. Aus technischen Gründen muß $g(\cdot)$ auf E streng monoton und differenzierbar sein.

2.2.2 Das dynamische Trendmodell

Betrachten wir nun ein Modell, das eine systematische Veränderung zwischen μ_{t-1} und μ_t zuläßt. Für normalverteilte Prozesse kann folgende dynamische stochastische Gleichung zur Modellierung der Veränderung von μ_t herangezogen werden, die ein Spezialfall des dynamischen Trendmodells ([14]) ist:

$$\mu_t = \mu_{t-1} + a + w_t, \quad w_t \sim N(0, Q).$$

In dieser Form besitzt das Modell eine "zeitinvariante Trendkomponente", da a nicht von der Zeit abhängt. Das dynamische Trendmodell in seiner allgemeinen Form ([14])

entsteht, wenn man annimmt, daß sich auch die Trendkomponente a_t zufällig mit der Zeit ändert:

$$\begin{aligned}\mu_t &= \mu_{t-1} + a_t + w_t, & w_t &\sim N(0, Q), \\ a_t &= a_{t-1} + \tilde{w}_t, & \tilde{w}_t &\sim N(0, W).\end{aligned}\tag{5}$$

Für nicht-normalverteilte Prozesse muß μ_t wieder entsprechend transformiert werden.

2.2.3 Die allgemeine Modellformulierung

Die Modelle der beiden vorangegangenen Abschnitte sind Spezialfälle einer wesentlich allgemeineren Modellform. Zur Motivation der allgemeinen Modellform betrachten wir nochmals das dynamische Trendmodell in seiner allgemeinen Form (5). Es enthält zwei Größen, die sich dynamisch ändern, nämlich den Erwartungswert μ_t und die Trendkomponente a_t . Die dynamisch sich verändernden Größen werden in einem Vektor zusammengefaßt, dem sogenannten Zustandsvektor \mathbf{x}_t . Der Zustandsvektor wird so gewählt, daß der Erwartungswert $\mu_t(\mathbf{x}_t)$ bei bekanntem \mathbf{x}_t deterministisch ist, z.B.:

	Zustandsvektor \mathbf{x}_t	$\mu_t(\mathbf{x}_t)$
Steady-State-Modell	$g(\mu_t)$	$g^{-1}(\mathbf{x}_t)$
Dynamisches Trendmodell	$\begin{pmatrix} g^{-1}(\mu_t) \\ a_t \end{pmatrix}$	$g^{-1}((1 \ 0) \cdot \mathbf{x}_t)$

Wir formulieren nun das dynamische stochastische Modell in seiner allgemeinen Form, die im wesentlichen auf Ideen in [14] und [33] zurückgeht. Ein dynamisches stochastisches Modell mit Zustandsvektor \mathbf{x}_t wird für jeden Zeitpunkt t durch zwei stochastische Gleichungen definiert. Die erste stochastische Gleichung ist dynamisch und beschreibt, wie sich der Zustandsvektor \mathbf{x}_t mit der Zeit verändert:

$$\mathbf{x}_t = \mathbf{F}_t \cdot \mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{Q}_t).\tag{6}$$

Diese Gleichung ist äquivalent mit der Angabe der Verteilung von \mathbf{x}_t gegeben \mathbf{x}_{t-1} (in Zeichen $\mathbf{x}_t|\mathbf{x}_{t-1}$):

$$\mathbf{x}_t|\mathbf{x}_{t-1} \sim N(\mathbf{F}_t \cdot \mathbf{x}_{t-1}, \mathbf{Q}_t).$$

Die zweite stochastische Gleichung beschreibt, wie der Prozeß y_t zum Zeitpunkt t verteilt ist, wenn der Zustandsvektor \mathbf{x}_t bekannt ist (in Zeichen $y_t|\mathbf{x}_t$). Dabei wird angenommen, daß der Erwartungswert μ_t der Beobachtungsverteilung von y_t – eventuell nach einer Transformation – linear mit dem Zustandsvektor \mathbf{x}_t zusammenhängt:

$$E(y_t|\mathbf{x}_t) = E(y_t|\mu_t(\mathbf{x}_t)) = \mu_t(\mathbf{x}_t) = g^{-1}(\mathbf{H}_t \cdot \mathbf{x}_t).\tag{7}$$

Für die Varianz gilt wegen (1):

$$V(y_t|\mathbf{x}_t) = V(y_t|\mu_t(\mathbf{x}_t)) = v_t(\mu_t(\mathbf{x}_t)) \cdot \phi.$$

Dynamische stochastische Modelle haben zahlreiche Anwendungen – allerdings im allgemeinen eingeschränkt auf normalverteilte Prozesse – gefunden, von denen wir nur eine kleine Auswahl aus dem Bereich der Ökologie und der Biometrie erwähnen können: zur Wasserqualitätskontrolle z.B. in [4] und [5], zur Modellierung hydrologischer und hydraulischer Prozesse z.B. in [6], [12] und [25], zur Luftqualitätskontrolle z.B. in [19], zum Monitoring von Nierentransplantationen in [22].

In der Praxis stellt sich die Frage, wie für einen konkreten Datensatz die Modellstruktur, d.h. der Zustandsvektor und die Modellmatrizen zu wählen sind. Der systemtheoretische Ansatz, der auf Realisierungen eines stationären normalverteilten Prozesses anwendbar ist, schätzt unter der Annahme zeitinvarianter Matrizen ein kanonisches Modell mit minimaler Dimension des Zustandsvektors ([2]). Die angewandte dynamische Modellierung geht eher von einem strukturellen Ansatz aus, bei dem in die Wahl des Zustandsvektors und der Modellmatrizen apriori vorhandene Vorstellungen über die den Prozeß verursachende Dynamik einfließen. Dieser Ansatz ist nicht auf Realisierungen stationärer Prozesse beschränkt und kann substanzwissenschaftliche Modellvorstellungen in den Modellidentifikationsprozeß einbinden (vgl. z.B. [30], [29], [10]).

Selbst wenn die prinzipielle Modellstruktur feststeht, verbleiben im allgemeinen Parameter, die noch zu spezifizieren sind. Ein Steady-State-Modell für normalverteilte Prozesse ist voll spezifiziert, wenn für die Varianzen Q und R konkrete Werte gewählt wurden. In ähnlicher Weise hängt auch das dynamische Trendmodell von Parametern ab, die in einem Vektors θ unter der Bezeichnung Hyperparameter zusammengefaßt werden. Die Komponenten dieses Hyperparameters sind in der Praxis meistens apriori unbekannt. Wir werden in Abschnitt 4 auf dieses Problem näher eingehen. Für Abschnitt 3 nehmen wir zunächst an, daß das Modell voll spezifiziert wurde, indem für den Hyperparameter konkrete Werte eingesetzt wurden.

2.3 Dynamische stochastische Modellierung von Datensatz 1 und 2

2.3.1 Datensatz 1

Da bei dieser Zeitreihe der Prozeß y_t durch Mittelung des Grundwasserspiegels über ein ganzes Jahr entsteht, können wir wegen des zentralen Grenzwertsatzes nehmen, daß y_t normalverteilt ist. Wir wählen daher folgende Beobachtungsverteilung:

$$y_t|\mu_t \sim N(\mu_t, R), \quad p(y_t|\mu_t) = \frac{1}{\sqrt{2\pi R}} \cdot \exp\left(-\frac{(y_t - \mu_t)^2}{2R}\right).$$

Die Varianz R von y_t um μ_t ist nach dieser Annahme zeitinvariant. Diese Beobachtungsverteilung erfüllt mit $v_t(\mu_t) = 1$ und $\phi = R$ Voraussetzung (1).

Da die Abbildung 1 deutlich zeigt, daß sich der Erwartungswert μ_t der Beobachtungen systematisch verändert hat, modellieren wir diese Daten mit dem dynamischen Trendmodell aus Abschnitt 2.2.2:

Modell 1 - Dynamisches Trendmodell für normalverteilte Prozesse ([14]): Der Zustandsvektor \mathbf{x}_t besteht aus zwei Komponenten:

$$\mathbf{x}_t = \begin{pmatrix} \mu_t \\ a_t \end{pmatrix},$$

wobei μ_t den Level des Prozesses y_t und a_t die systematische Veränderung des Levels, die sogenannte Trendkomponente, bezeichnet. Fassen wir das Modell in der Schreibweise der allgemeinen Modellform aus Abschnitt 2.2.3 zusammen:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(0, \mathbf{Q}), \\ \mathbf{F} &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} Q + W & W \\ W & W \end{pmatrix}, \\ y_t | \mu_t &\sim N(\mu_t, R), \\ \mu_t &= \mathbf{H}\mathbf{x}_t, \quad \mathbf{H} = \begin{pmatrix} 1 & 0 \end{pmatrix}. \end{aligned}$$

Der Hyperparameter θ umfaßt die Varianzen Q , W und R . Diese Varianzen sind bei der Anwendung des Modells auf den Datensatz 1 apriori unbekannt.

2.3.2 Datensatz 2

Abbildung 2 zeigt deutlich, daß die Verteilung des Prozesses „Tagesmittelwert von SO₂-Emissionen“ trotz Mittelbildung schief ist. Weiters nimmt die Varianz der Beobachtungen mit wachsendem Erwartungswert zu. Wir wählen deshalb zur Modellierung dieses Prozesses eine Gamma-Verteilung:

$$y_t | \mu_t \sim \gamma\left(\alpha, \frac{\alpha}{\mu_t}\right), \quad p(y_t | \mu_t) = \frac{1}{\Gamma(\alpha)} \left(\frac{\alpha}{\mu_t}\right)^\alpha y_t^{\alpha-1} \exp\left(-\frac{\alpha \cdot y_t}{\mu_t}\right).$$

Die Schiefe dieser Verteilung wird durch den Parameter α gesteuert. Für $\alpha = 1$ erhalten wir einen exponentialverteilten Prozeß. Mit wachsendem α nähert sich der Prozeß einem normalverteilten Prozeß mit zeitvarianter Varianz. Da die Varianz von y_t um μ_t bei einer Gamma-Verteilung proportional zum Quadrat des Erwartungswertes μ_t ist:

$$V(y_t | \mu_t) = \frac{\mu_t^2}{\alpha},$$

modelliert diese Verteilung neben der Schiefe auch die Inhomogenität der Varianz der Daten. Diese Beobachtungsverteilung erfüllt Voraussetzung (1) mit

$$v_t(\mu_t) = \mu_t^2, \quad \phi = \frac{1}{\alpha}.$$

Abbildung 2 zeigt keine systematische Veränderung des Erwartungswertes. Da der Erwartungswert andererseits auch nicht konstant zu sein scheint, modellieren wir die Daten mit folgendem Steady-State-Modell für Gamma-verteilte Prozesse.

Modell 2 Steady-State-Modell für Gamma-verteilte Prozesse

Der Zustandsvektor ist eindimensional und mit dem über $g(\cdot)$ transformierten Level μ_t des Prozesses y_t identisch. Wir definieren die Transformation $g(\mu)$ über:

$$g(\mu) = \begin{cases} \ln \mu + 1, & \mu \leq 1, \\ \mu & \mu \geq 1. \end{cases}$$

Bei dieser Transformation wird für den Bereich $\mu < 1$ mit unsymmetrischen und für den Bereich $\mu \geq 1$ mit symmetrischen Schwankungen gearbeitet.

Das Modell ist durch die folgenden stochastischen Gleichungen definiert:

$$\begin{aligned} x_t &= x_{t-1} + w_t, & w_t &\sim N(0, Q), \\ y_t | \mu_t &\sim \gamma\left(\alpha, \frac{\alpha}{\mu_t}\right), \\ \mu_t &= g^{-1}(x_t). \end{aligned}$$

Das Modell besitzt die allgemeine Modellform aus Abschnitt 2.2.3 mit $F_t = 1$ und $H_t = 1$. Der Hyperparameter θ umfaßt die Varianz Q und den Parameter α der Beobachtungsverteilung. Beide Werte sind bei der Anwendung des Modells auf den Datensatz 2 apriori unbekannt.

3 Aussagen über unbeobachtbare Größen

3.1 Allgemeine Bemerkungen

Wir gehen nun von einem Prozeß aus, der bis zum Zeitpunkt t beobachtet wurde. Die Zeitreihe y_1, \dots, y_t der Beobachtungswerte bis t wird mit y^t abgekürzt. Aufbauend auf den Beobachtungen y^t interessieren nun statistische Aussagen über zum Zeitpunkt t unbeobachtbare Größen u (in Zeichen $u|y^t$). Unter einer zum Zeitpunkt t unbeobachtbaren Größe verstehen wir eine Größe, die stochastisch ist, wenn Beobachtungen bis zum Zeitpunkt t vorliegen. Unbeobachtbare Größen sind der Erwartungswert $\mu_s|y^t$ oder der Zustandsvektor $\mathbf{x}_s|y^t$ zu jedem beliebigen Zeitpunkt s sowie zukünftige Werte des Prozesses $y_s|y^t, s > t$.

Bei der Beobachtung von stochastischen Prozessen entsteht durch die Zeit eine Ordnung in den Beobachtungswerten. Wird der Zeitpunkt t als Gegenwart ausgezeichnet, so entsteht automatisch Vergangenheit und Zukunft. Bei Aussagen über unbeobachtbare Größe können daher Aussagen über die Gegenwart, die Vergangenheit und die Zukunft unterschieden werden. Aussagen über $\mathbf{x}_t|y^t$ und $\mu_t|y^t$ sind Aussagen über die Gegenwart. Aussagen über $\mathbf{x}_s|y^t$ und $\mu_s|y^t$ mit $s < t$ sind Aussagen über die Vergangenheit, Aussagen über $\mathbf{x}_s|y^t, \mu_s|y^t$ und $y_s|y^t$ mit $s > t$ sind Aussagen über die Zukunft.

Die stochastische Variation der unbeobachtbaren Größe $u|y^t$ ist durch eine Verteilungsfunktion bzw. deren Dichte $p(u|y^t)$ beschreibbar. Aus der Dichte läßt sich die Wahrscheinlichkeit berechnen, mit der eine Aussage über $u|y^t$ zutrifft. Eine umfassende Lösung des statistischen Inferenzproblems besteht in der Bestimmung der Dichten $p(u|y^t)$ für alle interessierenden Größen $u|y^t$. Die Dichte $p(\mathbf{x}_t|y^t)$ des gegenwärtigen Zustandsvektors $\mathbf{x}_t|y^t$ wird Filterdichte genannt. Die Dichte $p(\mathbf{x}_s|y^t)$ eines vergangenen Zustandsvektors $\mathbf{x}_s|y^t, s < t$, heißt Glättungsdichte. Die Dichte $p(\mathbf{x}_s|y^t)$ des zukünftigen Zustandsvektors $\mathbf{x}_s|y^t$ und die Dichte $p(y_s|y^t)$ eines zukünftigen Prozeßwertes $y_s|y^t, s > t$, heißen Vorhersagedichten.

Liegt eine Datenreihe fixer Länge N zur Analyse vor, so spricht man von einer off-line-Analyse (z.B. Trendanalyse der Grundwasserdaten). Treffen hingegen laufend neue Beobachtungen ein, so verschiebt sich der Zeitpunkt der Analyse laufend und man spricht von einer on-line Analyse (z.B. laufendes Monitoring der SO_2 -Konzentrationen). Im zweiten Fall ist der als Gegenwart ausgezeichnete Zeitpunkt t mit dem Analysezeitpunkt identisch. Kommt eine neue Beobachtung hinzu, so verschiebt sich die Gegenwart um eine Zeiteinheit. Es stellt sich dann die Frage, wie Aussagen zum Zeitpunkt t mit den Aussagen zum Zeitpunkt $t - 1$ zusammenhängen. Eine Prozedur, die angibt, wie die Filterdichten aufeinanderfolgender Zeitschritte zusammenhängen, wird als statistischer Filter bezeichnet (siehe Abschnitt 3.2).

Im Falle der off-line-Analyse interessieren vorwiegend die Glättungsdichten $p(\mathbf{x}_s|y^N)$, $s \leq N$, für den letzten Zeitpunkt N . Die Bestimmung dieser Dichten erfolgt auch bei der

off-line-Analyse am einfachsten, indem zunächst eine on-line-Analyse mit laufender Verschiebung der Gegenwart von $t = 1$ bis $t = N$ durchgeführt und die Filterdichten $p(\mathbf{x}_t|y^t)$ ermittelt werden (siehe Abschnitt 3.2). In ähnlicher Weise baut auch die Ermittlung der Vorhersagedichten $p(\mathbf{x}_s|y^t)$ und $p(y_s|y^t)$, $s > t$, auf der Filterdichte $p(\mathbf{x}_t|y^t)$ auf (siehe Abschnitt 3.4).

3.2 Filter- und Glättungsdichten

Wird ein vollspezifiziertes dynamisches stochastisches Modell auf einen normalverteilten Prozeß angewendet, so sind die Filterdichten Dichten einer Normalverteilung und durch die beiden ersten Momente charakterisiert. Zwischen den Momenten der Filterdichte $p(\mathbf{x}_{t-1}|y^{t-1})$ und den Momenten der Filterdichte $p(\mathbf{x}_t|y^t)$ besteht folgender linearer Zusammenhang (Kalman-Filter, [15]):

$$\boxed{\mathbf{x}_{t-1}|y^{t-1} \sim N(\hat{\mathbf{x}}_{t-1|t-1}, \mathbf{P}_{t-1|t-1}) \Rightarrow \mathbf{x}_t|y^t \sim N(\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})} \quad (8)$$

$$\begin{aligned} \hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(y_t - \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1}), \\ \mathbf{P}_{t|t} &= (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1}, \quad \mathbf{I} \dots \text{Einheitsmatrix,} \\ \mathbf{K}_t &= \mathbf{P}_{t|t-1}\mathbf{H}_t^T (\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^T + \mathbf{R}_t)^{-1}, \\ \hat{\mathbf{x}}_{t|t-1} &= \mathbf{F}_t\hat{\mathbf{x}}_{t-1|t-1}, \\ \mathbf{P}_{t|t-1} &= \mathbf{F}_t\mathbf{P}_{t-1|t-1}\mathbf{F}_t^T + \mathbf{Q}_t. \end{aligned} \quad (9)$$

Beachtenswert ist, daß in die Ermittlung des Zusammenhangs zwischen den ersten Momenten aufeinanderfolgender Filterdichten nur die aktuelle Beobachtung y_t einfließt. Die Filterdichte zum Zeitpunkt $t - 1$ enthält alle Information der Daten bis zum Zeitpunkt $t - 1$. Die Kovarianzmatrix der Filterdichte hängt nicht von den Beobachtungen ab.

Das sequentielle Schema zeigt, daß man zur Ermittlung der Filterdichte $p(\mathbf{x}_1|y^1)$ zum Zeitpunkt $t = 1$ eine a-priori-Dichte $p(\mathbf{x}_0|y^0)$ vorgeben muß. Sie kann durch entsprechende Wahl der Parameter als nicht informativ angenommen werden (siehe Abschnitt 3.3.1).

Für nicht-normalverteilte Prozesse kann ein dem Kalman-Filter vergleichbares sequentielles Schema nur direkt für die Filterdichten unter Anwendung des Bayes'schen Theorems abgeleitet werden (z.B. [32]):

$$\begin{aligned} p(\mathbf{x}_t|y^t) &\propto p(y_t|\mathbf{x}_t) \cdot p(\mathbf{x}_t|y^{t-1}), \\ p(\mathbf{x}_t|y^{t-1}) &= \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|y^{t-1}) d\mathbf{x}_{t-1}. \end{aligned} \quad (10)$$

Für normalverteilte Prozesse läßt sich daraus ein sequentielles Schema für die Momente der Filterdichten, eben der Kalman-Filter, herleiten. Für nicht-normalverteilte Prozesse existiert im allgemeinen kein exaktes sequentielles Schema für die Momente.

Den in der Literatur vorgeschlagenen approximativen Filtern ([33], [9], [27]) ist gemeinsam, daß sie keine Aussage über die Gestalt der gesamten Filterdichte, sondern nur

über gewisse Charakteristika wie Lage oder Streuung ermöglichen. Eine Approximation der gesamten Filterdichte wurde von [16] mit extremen numerischen Aufwand versucht. Ähnliche Ansätze sind auch in [32] zu finden. In den Abschnitten 3.2.1 und 3.2.2 beschreiben wir im Detail zwei Filter, die sich im Rahmen von Simulationsstudien für ein dynamisches Trendmodell ([27]) bewährt haben.

Wenden wir uns nun den Glättungsdichten zu. Wird ein vollspezifiziertes dynamisches stochastisches Modell auf einen normalverteilten Prozeß angewendet, so ist die Glättungsdichte $p(\mathbf{x}_s|y^t)$, $s < t$, die Dichte einer Normalverteilung, deren Momente aus den Momenten der Filterdichte $p(\mathbf{x}_s|y^s)$ und den Momenten der Glättungsdichte $p(\mathbf{x}_{s+1}|y^t)$ bestimmt werden können (vgl. z.B. [32]):

$$\boxed{\mathbf{x}_s|y^s \sim N(\hat{\mathbf{x}}_{s|s}, \mathbf{P}_{s|s}), \mathbf{x}_{s+1}|y^t \sim N(\hat{\mathbf{x}}_{s+1|t}, \mathbf{P}_{s+1|t}) \Rightarrow \mathbf{x}_s|y^t \sim N(\hat{\mathbf{x}}_{s|t}, \mathbf{P}_{s|t})} \quad (11)$$

$$\begin{aligned} \hat{\mathbf{x}}_{s|t} &= \hat{\mathbf{x}}_{s|s} + \mathbf{A}_{s+1}(\hat{\mathbf{x}}_{s+1|t} - \mathbf{F}_{s+1}\hat{\mathbf{x}}_{s|s}), \\ \mathbf{P}_{s|t} &= \mathbf{P}_{s|s} + \mathbf{A}_{s+1}(\mathbf{P}_{s+1|t} - \mathbf{P}_{s+1|s})\mathbf{A}_{s+1}^T, \\ \mathbf{P}_{s+1|s} &= \mathbf{F}_{s+1}\mathbf{P}_{s|s}\mathbf{F}_{s+1}^T + \mathbf{Q}_{s+1}, \\ \mathbf{A}_{s+1} &= \mathbf{P}_{s|s}\mathbf{F}_{s+1}^T(\mathbf{P}_{s+1|s})^{-1}. \end{aligned}$$

Für nicht-normalverteilte Prozesse ist die Glättungsdichte keine Normalverteilung. Ein approximatives Schema für die beiden ersten Momente der Glättungsdichte wurde von [9] abgeleitet, das genau obige Form besitzt. $\hat{\mathbf{x}}_{s|s}$ und $\mathbf{P}_{s|s}$ sind die beiden ersten Momente der Filterdichte, die durch die Charakteristika des approximativen Filters angenähert werden.

3.2.1 Approximativer Posterior-Mode-Filter

Dieser Filter wurde von [9] als approximativer Filter für den Modus und die Inverse der Informationsmatrix am Modus der Filterdichte abgeleitet. Derselbe Filter entsteht als approximativer Filter für die beiden ersten Momente der Filterdichte, wenn man für jeden Zeitpunkt t das nichtlineare Modell lokal linearisiert. Für normalverteilte Prozesse kann aus dem sequentiellen Schema (10) für die Dichten deshalb ein sequentielles Schema für die Momente abgeleitet werden, weil der Erwartungswert von $y_t|\mathbf{x}_t$ gegeben \mathbf{x}_t linear in \mathbf{x}_t ist und die Varianz von $y_t|\mathbf{x}_t$ von \mathbf{x}_t nicht abhängt. Für nicht-normalverteilte Prozesse sind beide Voraussetzungen nicht erfüllt (siehe Abschnitt 2.1):

$$\begin{aligned} E(y_t|\mathbf{x}_t) &= \mu_t(\mathbf{x}_t) = g^{-1}(\mathbf{H}_t\mathbf{x}_t) \\ V(y_t|\mathbf{x}_t) &= \phi \cdot v(\mu_t(\mathbf{x}_t)) = \phi \cdot v(g^{-1}(\mathbf{H}_t\mathbf{x}_t)). \end{aligned}$$

Ein approximatives sequentielles Schema für die Momente kann abgeleitet werden, wenn der Erwartungswert $E(y_t|\mathbf{x}_t)$ in \mathbf{x}_t lokal um den bedingten Erwartungswert $\mathbf{x}_t^0 = E(\mathbf{x}_t|y^{t-1})$ linearisiert wird und die Varianz $V(y_t|\mathbf{x}_t)$ durch einen von \mathbf{x}_t unabhängigen Wert angenähert wird:

$$E(y_t|\mathbf{x}_t) \approx g^{-1}(\mathbf{H}_t\mathbf{x}_t^0) + \mathbf{H}_t^*(\mathbf{x}_t^0) \cdot (\mathbf{x}_t - \mathbf{x}_t^0),$$

$$\begin{aligned} \mathbf{H}_t^*(\mathbf{x}_t^0) &= (g^{-1})'(\mathbf{H}_t \mathbf{x}_t^0) \cdot \mathbf{H}_t, \\ V(y_t | \mathbf{x}_t) &\approx \phi \cdot v_t(g^{-1}(\mathbf{H}_t \mathbf{x}_t^0)) =: R_t^*(\mathbf{x}_t^0). \end{aligned}$$

Der Filter, der sich durch diese lokale Linearisierung ergibt, hat eine Form, die dem Kalman-Filter ähnlich ist ([9]):

$$\boxed{\mathbf{x}_{t-1} | y^{t-1} \sim (\hat{\mathbf{x}}_{t-1|t-1}^F, \hat{\mathbf{P}}_{t-1|t-1}^F) \Rightarrow \mathbf{x}_t | y^t \sim (\hat{\mathbf{x}}_{t|t}^F, \hat{\mathbf{P}}_{t|t}^F)} \quad (12)$$

$$\begin{aligned} \hat{\mathbf{x}}_{t|t}^F &= \mathbf{x}_t^0 + \mathbf{K}_t(\mathbf{x}_t^0)(y_t - g^{-1}(\mathbf{H}_t \mathbf{x}_t^0)), \\ \hat{\mathbf{P}}_{t|t}^F &= (\mathbf{I} - \mathbf{K}_t(\mathbf{x}_t^0) \mathbf{H}_t^*(\mathbf{x}_t^0)) \mathbf{P}_{t|t-1}, \\ \mathbf{x}_t^0 &= \mathbf{F}_t \hat{\mathbf{x}}_{t-1|t-1}^F, \\ \mathbf{K}_t(\mathbf{x}_t^0) &= \mathbf{P}_{t|t-1} (\mathbf{H}_t^*(\mathbf{x}_t^0))^T \left(\mathbf{H}_t^*(\mathbf{x}_t^0) \mathbf{P}_{t|t-1} (\mathbf{H}_t^*(\mathbf{x}_t^0))^T + R_t^*(\mathbf{x}_t^0) \right)^{-1}, \\ \mathbf{P}_{t|t-1} &= \mathbf{F}_t \hat{\mathbf{P}}_{t-1|t-1}^F \mathbf{F}_t^T + \mathbf{Q}_t. \end{aligned}$$

Für normalverteilte Prozesse ist dieser Filter mit dem Kalman-Filter identisch. Gilt bei nicht-normalverteilten Prozessen $g^{-1}(\mathbf{H}_t \mathbf{x}_t^0) = \mathbf{H}_t \mathbf{x}_t^0$, so hat dieser Filter zwar die Form eines Kalman-Filters mit Beobachtungsvarianz $R_t = \phi \cdot v_t(\mathbf{H}_t(\mathbf{x}_t^0))$, der Zusammenhang zwischen den beiden Lagecharakteristika bleibt aber wegen der Abhängigkeit der Beobachtungsvarianz von $\hat{\mathbf{x}}_{t-1|t-1}^F$ nichtlinear. Ein weiterer Unterschied zum Kalman-Filter ist die Abhängigkeit der Streuungscharakteristika der Filterdichte von den Beobachtungen y^t über $\hat{\mathbf{x}}_{t-1|t-1}^F$.

3.2.2 Filter auf Basis orthogonaler Integration

Dieser Filter wurde in [27] vorgeschlagen und am Beispiel eines verallgemeinerten dynamischen Trendmodells für Gamma- und Poisson-verteilte Prozesse illustriert. Die ersten zwei Momente der Filterdichte $p(\mathbf{x}_t | y^t)$ werden sequentiell aus den Momenten $\hat{\mathbf{x}}_{t-1|t-1}^I$ und $\hat{\mathbf{P}}_{t-1|t-1}^I$ der Filterdichte zum Zeitpunkt $t-1$ ermittelt, indem die nicht normierte Filterdichte $p^*(\mathbf{x}_t | y^t) = p(y_t | \mathbf{x}_t) \cdot p(\mathbf{x}_t | y^{t-1})$, die sich aus dem Bayes'schen Theorem (10) ergibt, numerisch integriert wird:

$$\hat{\mathbf{x}}_{t|t}^I = E(\mathbf{x}_t | y^t) = \frac{1}{C_t} \int \mathbf{x}_t p^*(\mathbf{x}_t | y^t) d\mathbf{x}_t, \quad (13)$$

$$\hat{\mathbf{P}}_{t|t}^I = V(\mathbf{x}_t | y^t) = \frac{1}{C_t} \int \mathbf{x}_t \mathbf{x}_t^T p^*(\mathbf{x}_t | y^t) d\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}^I (\hat{\mathbf{x}}_{t|t}^I)^T, \quad (14)$$

$$C_t = \int p^*(\mathbf{x}_t | y^t) d\mathbf{x}_t. \quad (15)$$

Die Integration über den r -dimensionalen Zustandsvektor \mathbf{x}_t wird für jeden Zeitpunkt t mittels multivariater Gauß-Hermite-Integration durchgeführt, einer Integrationsmethode, die sich in der Bayes'schen Analyse häufig bewährt hat (z.B. [23], [18], [25]).

Die Stützstellen $\mathbf{x}_t^{(i)}$ mit den Gewichten $w_I^{(i)}$ entstehen durch Transformation eines cartesischen Gitters in \mathbb{R}^r :

$$\begin{aligned}\mathbf{x}_t^{(i)} &= \mathbf{m}_t + \mathbf{U}_t \cdot \boldsymbol{\tau}^{(i)}, & \mathbf{U}_t \mathbf{U}_t^T &= 2\mathbf{S}_t, \\ \boldsymbol{\tau}^{(i)} &= \begin{pmatrix} \tau^{(i_1)} \\ \vdots \\ \tau^{(i_r)} \end{pmatrix}, & w_I^{(i)} &= \left(\frac{1}{\sqrt{\pi}} \right)^r \cdot \omega^{(i_1)} \dots \omega^{(i_r)}, \quad 1 \leq i_1, \dots, i_r \leq M_I.\end{aligned}$$

$\tau^{(1)}, \dots, \tau^{(M_I)}$ sind die Nullstellen eines Hermite-Polynoms vom Grade M_I , $\omega^{(1)}, \dots, \omega^{(M_I)}$ sind die Integrationsgewichte einer Gauß-Hermite-Integration mit Gewichtsfunktion $\exp(-x^2)$ ([1], S. 924, Tabelle 25.10). Die Transformationsparameter \mathbf{m}_t und \mathbf{S}_t werden so gewählt, daß das Gitter in einen Bereich fällt, über dem sich der Integrand, das heißt die unnormierte a-posteriori-Dichte, mit Lage \mathbf{m}_t und Streuung \mathbf{S}_t , konzentriert. Wir wählen daher als Transformationsparameter \mathbf{m}_t und \mathbf{S}_t jene Charakteristika der a-posteriori-Dichte, die durch den approximativen Posterior-Mode-Filter (Abschnitt 3.2.1) berechnet wurden, wobei statt $\hat{\mathbf{x}}_{t-1|t-1}^F$ und $\hat{\mathbf{P}}_{t-1|t-1}^F$ die Momente $\hat{\mathbf{x}}_{t-1|t-1}^I$ und $\hat{\mathbf{P}}_{t-1|t-1}^I$ der Filterdichte zum Zeitpunkt $t-1$ eingesetzt werden. Der Filter, der auf diese Weise aus der Integration von (13) – (15) entsteht, läßt sich als Korrektor des Posterior-Mode-Filters darstellen:

$$\begin{aligned}\hat{\mathbf{x}}_{t|t}^I &= \hat{\mathbf{x}}_{t|t}^F + \mathbf{U}_t \cdot \mathbf{z}_t, \\ \hat{\mathbf{P}}_{t|t}^I &= \mathbf{U}_t \mathbf{Z}_t \mathbf{U}_t^T, & \hat{\mathbf{P}}_{t|t}^F &= 0.5 \mathbf{U}_t \mathbf{U}_t^T, \\ \mathbf{z}_t &= \sum_{i=1}^{(M_I)^r} \psi(\mathbf{x}_t^{(i)}), \\ \mathbf{z}_t &= \frac{1}{z_t} \sum_{i=1}^{(M_I)^r} \boldsymbol{\tau}^{(i)} \psi(\mathbf{x}_t^{(i)}), \\ \mathbf{Z}_t &= \frac{1}{z_t} \sum_{i=1}^{(M_I)^r} \boldsymbol{\tau}^{(i)} (\boldsymbol{\tau}^{(i)})^T \psi(\mathbf{x}_t^{(i)}) - \mathbf{z}_t \mathbf{z}_t^T, \\ \psi(\mathbf{x}_t^{(i)}) &= \frac{p(y_t | \mathbf{x}_t^{(i)}) p_N(\mathbf{x}_t^{(i)}; \mathbf{x}_t^0, \mathbf{P}_{t|t-1})}{p_N(\mathbf{x}_t^{(i)}; \hat{\mathbf{x}}_{t|t}^F, \hat{\mathbf{P}}_{t|t}^F)} w_I^{(i)}.\end{aligned} \tag{16}$$

$p_N(\mathbf{x}_t^{(i)}; \mathbf{x}_t^0, \mathbf{P}_{t|t-1})$ bezeichnet den Funktionswert der Dichte einer $N(\mathbf{x}_t^0, \mathbf{P}_{t|t-1})$ -Verteilung an der Stelle $\mathbf{x}_t^{(i)}$. Für normalverteilte Prozesse ergibt diese Approximation mit $M_I \geq 2$ den *exakten* Filter, da in diesem Fall $\mathbf{z}_t = \mathbf{0}$, $\mathbf{Z}_t = \frac{1}{2} \mathbf{I}$ und der Posterior-Mode-Filter mit dem Kalman-Filter identisch ist.

3.3 Fallstudien zum Filterproblem

3.3.1 Datensatz 1: Fallstudie 1

Analysieren wir, welche Aussagen über die Veränderung des Jahresmittelwertes des Grundwasserspiegels möglich sind, wenn wir den Datensatz 1 mit einem dynamischen Trendmodell für normalverteilte Prozesse (Modell 1) beschreiben.

Kalman-Filter für das dynamische Trendmodell. Wir diskutieren zunächst das Schema, das sich aus dem Kalman-Filter zur sequentiellen Ermittlung der ersten Momente der Filterdichte des Levels $\mu_t|y^t$ und der Trendkomponente $a_t|y^t$ ergibt. Mit den Bezeichnungen:

$$\hat{\mu}_t = E(\mu_t|y^t), \quad \hat{a}_t = E(a_t|y^t),$$

erhalten wir aus (8) folgenden Zusammenhang zwischen den ersten Momenten der Filterdichten aufeinanderfolgender Zeitpunkte:

$$\begin{aligned}\hat{\mu}_t &= (1 - \mathbf{K}_{t,1})(\hat{\mu}_{t-1} + \hat{a}_{t-1}) + \mathbf{K}_{t,1} \cdot y_t, \\ \hat{a}_t &= (1 - \mathbf{K}_{t,2})\hat{a}_{t-1} + \mathbf{K}_{t,2}(y_t - \hat{\mu}_{t-1}).\end{aligned}$$

$\hat{\mu}_{t-1} + \hat{a}_{t-1}$ ist eine Punktprognose für den Level zum Zeitpunkt t , wenn Beobachtungen bis $t - 1$ vorliegen. Liegt die Beobachtung y_t zum Zeitpunkt t vor, so ist das erste Moment des Levels zum Zeitpunkt t ein gewichtetes Mittel dieser Punktprognose und der tatsächlichen Beobachtung. Der Gewichtungsfaktor $\mathbf{K}_{t,1}$ ist die erste Komponente des Vektors \mathbf{K}_t in Gleichung (9):

$$\mathbf{K}_{t,1} = \frac{\|\mathbf{P}_{t-1|t-1}\|_S + Q + W}{\|\mathbf{P}_{t-1|t-1}\|_S + Q + W + R}.$$

$\|\mathbf{P}_{t-1|t-1}\|_S$ ist die Summe aller Elemente der Kovarianzmatrix $\mathbf{P}_{t-1|t-1}$. Dieser Gewichtungsfaktor liegt offenbar zwischen 0 und 1. Bei gleichbleibender Prozeßvarianz $Q + W$ ist das Gewicht der neuen Beobachtung umso größer, je kleiner die Beobachtungsvarianz R ist. Bei gleichbleibender Beobachtungsvarianz R ist dieses Gewicht umso größer, je größer die Prozeßvarianz $Q + W$ ist.

Auch das erste Moment der Trendkomponente ist ein gewichtetes Mittel aus dem ersten Moment zum Zeitpunkt $t - 1$ und der Größe $y_t - \hat{\mu}_{t-1}$, die einer indirekten Beobachtung der Trendkomponente entspricht. Der Gewichtungsfaktor $\mathbf{K}_{t,2}$ ist die zweite Komponente des Vektors \mathbf{K}_t in Gleichung (9):

$$\mathbf{K}_{t,2} = \frac{\mathbf{P}_{t-1|t-1,12} + \mathbf{P}_{t-1|t-1,22} + W}{\|\mathbf{P}_{t-1|t-1}\|_S + Q + W + R}.$$

$(\mathbf{P}_{t-1|t-1,12} + \mathbf{P}_{t-1|t-1,22})$ ist die Summe über die zweite Spalte der Kovarianzmatrix $\mathbf{P}_{t-1|t-1}$. Diese Gewichtung von Punktvorhersage und Beobachtung mit Gewichtungsfaktoren, die vom Verhältnis zwischen Beobachtungs- und Prozeßvarianz abhängen, ist charakteristisch für das Filterschema der ersten Momente bei normalverteilten Prozessen.

Für die Varianzen gilt folgender Zusammenhang:

$$P_{t|t} = \begin{pmatrix} K_{t,1}R & K_{t,1}R \\ K_{t,2}R & P_{t-1|t-1,22} + W - K_{t,2}^2(\|P_{t-1|t-1}\|_S + Q + W + R) \end{pmatrix}.$$

Es läßt sich zeigen (z.B. [3]), daß diese Matrix für $t \rightarrow \infty$ gegen eine Matrix konvergiert, die nur von Q , W und R abhängt.

Nicht-informative a-priori-Dichte. Wählen wir als a-priori-Dichte für den Level und die Trendkomponente die Dichte:

$$\hat{\mu}_0 = 0, \quad \hat{a}_0 = 0, \quad P_{0|0} = \begin{pmatrix} d^2 & 0 \\ 0 & d \end{pmatrix}, \quad (17)$$

mit sehr großem Wert d , so erhalten wir für $t = 1$ und $t = 2$ folgende Gewichtungsfaktoren und Momente der Filterdichte:

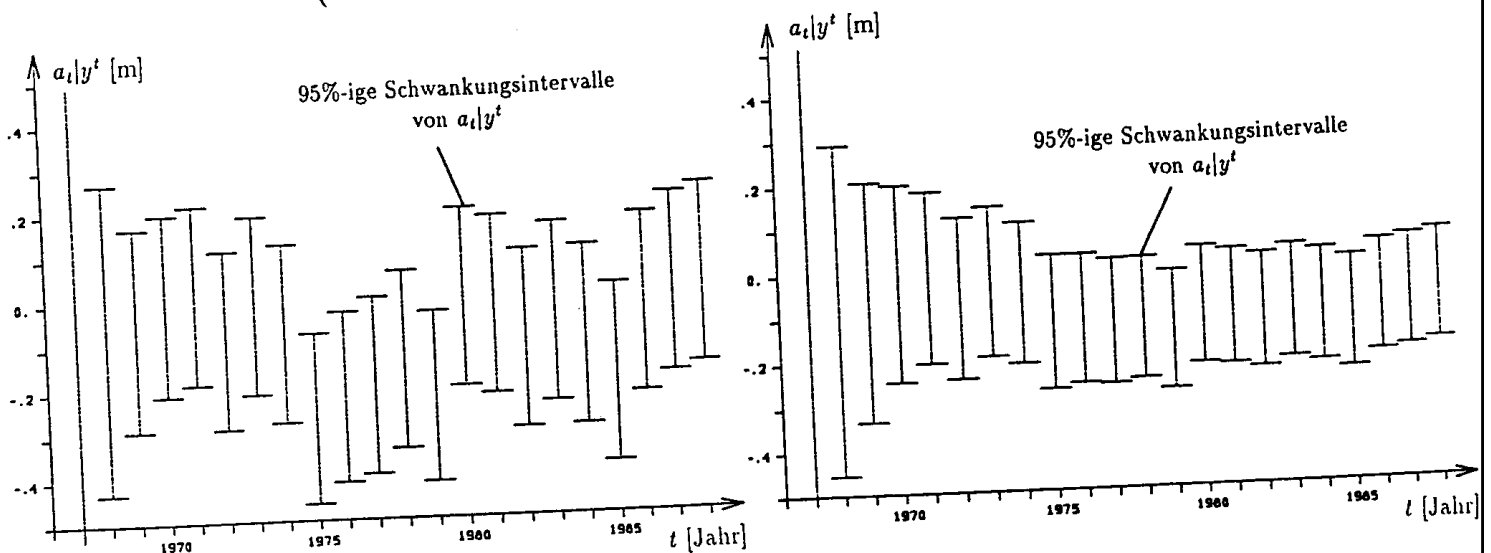
$$K_{1,1} \approx 1, \quad K_{1,2} \approx 0, \quad \hat{\mu}_1 \approx y_1, \quad \hat{a}_1 \approx 0, \quad P_{1|1} \approx \begin{pmatrix} R & 0 \\ 0 & d \end{pmatrix},$$

$$K_{2,1} \approx 1, \quad K_{2,2} \approx 1, \quad \hat{\mu}_2 \approx y_2, \quad \hat{a}_2 \approx y_1 - y_2.$$

Die Filterdichte von $\mu_1|y^1$ ist informativ, während die Filterdichte von $a_1|y^1$ uninformativ bleibt und erst zum Zeitpunkt $t = 2$ informativ wird.

Abbildung 3: Datensatz 1 - Fallstudie 1

95%-ige Schwankungsintervalle für die Trendkomponente $a_t|y^t$
(links: 1. Hyperparameter, rechts: 2. Hyperparameter)



Filtern für den Datensatz 1. Wir wenden das dynamische Trendmodell auf Datensatz 1 mit zwei verschiedenen Hyperparametern $\theta = (Q, W, R)$ an:

	Q	W	R
1. Hyperparameter	0.01	0.01	0.01
2. Hyperparameter	$0.341 \cdot 10^{-1}$	$0.343 \cdot 10^{-4}$	$0.409 \cdot 10^{-4}$

Der 1. Hyperparameter ist ein heuristisch gewählter Wert. Der 2. Hyperparameter wurde aus den Daten mit Methoden geschätzt, die wir in Abschnitt 4 diskutieren und in Abschnitt 4.5.1 auf den Datensatz 1 anwenden werden. Als Parameter der a-priori-Dichte (17) wurde $d = 10^4$ gewählt.

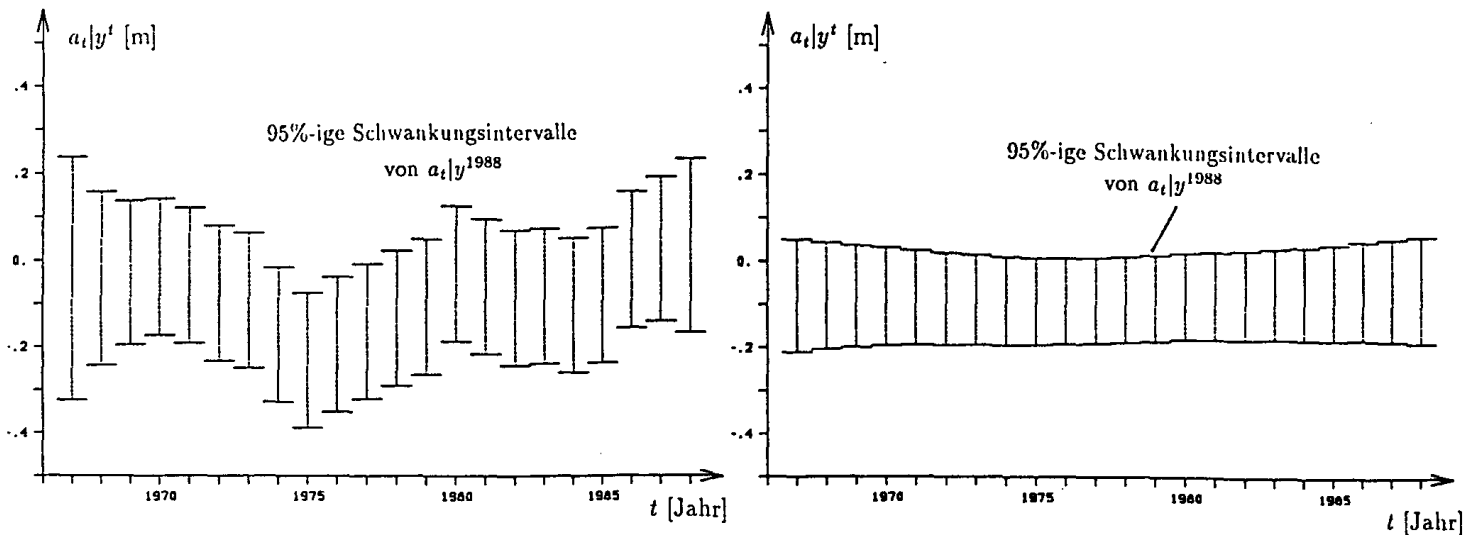
Wir vergleichen nun die Filterdichten der Trendkomponente $a_t|y^t$ für die beiden Hyperparameter. Für jeden Zeitpunkt t wurde aus der normalverteilten Filterdichte $p(a_t|y^t)$ ein 95%-iges Schwankungsintervall $S_{t|t}^{0.95}$ ermittelt:

$$S_{t|t}^{0.95} = [\hat{a}_t - 1.96\sqrt{P_{t|t,22}}, \hat{a}_t + 1.96\sqrt{P_{t|t,22}}].$$

$S_{t|t}^{0.95}$ ist ein Bereich, in den die Trendkomponente $a_t|y^t$ auf Grund der Beobachtungen bis zum Zeitpunkt t mit Wahrscheinlichkeit 0.95 fällt. In Abbildung 3 sind diese Schwankungsintervalle für beide Hyperparameter über t aufgetragen. Sie werden mit steigender Zahl der Beobachtungen schmaler. Ihre Breite hängt nur von der Varianz der Filterdichte ab und konvergiert wegen der Konvergenz der Varianzen gegen einen festen Wert, nämlich für den ersten Hyperparameter gegen 0.382 m und für den zweiten gegen 0.227 m. Die Grenzen der Intervalle konvergieren nicht, da sie vom ersten Moment der Filterdichte abhängen, das wegen der Abhängigkeit von y_t einem stochastischen Prozeß folgt.

Abbildung 4: Datensatz 1 - Fallstudie 1

95%-ige Schwankungsintervalle für die Trendkomponente $a_t|y^{1988}$
(links: 1. Hyperparameter, rechts: 2. Hyperparameter)



Aussagen über die systematische Veränderungen des Jahresmittelwertes des Grundwasserspiegels innerhalb eines Jahres. Berechnen wir nun für das Jahr t die Wahrscheinlichkeit, daß $a_t < 0$ war. Aus der Filterdichten können wir die Wahrscheinlichkeit $P(a_t < 0|y^t)$ aufbauend auf den Beobachtungen bis zum Jahre t berechnen.

Möchte man alle Beobachtungen einbeziehen, so müssen zuerst ausgehend von den Filterdichten die Glättungsdichten für $s = 1988, 1987, \dots, t$ nach dem Schema (11) ermittelt werden.

Zur Illustration sind in Abbildung 4 die Schwankungsintervalle $S_{t|1988}^{0.95}$ der Trendkomponente $a_t|y^{1988}$ dargestellt, die für die beiden verschiedenen Hyperparameter für jeden Zeitpunkt aus den Glättungsdichten $p(a_t|y^{1988})$ ermittelt wurden. Ein Vergleich mit den Schwankungsintervallen der Filterdichte zeigt deutlich den Einfluß, den die Anzahl der einbezogenen Beobachtungen auf die Breite von Schwankungsintervallen ausübt.

Tabelle 3: Datensatz 1 – Wahrscheinlichkeit $P(a_t < 0|y^{1988})$

für die verschiedenen Fallstudien

t	Fallstudie 1 (1.Hyperp.)	Fallstudie 1 (2.Hyperp.)	Fallstudie 2	t	Fallstudie 1 (1.Hyperp.)	Fallstudie 1 (2.Hyperp.)	Fallstudie 2
1967	0.6186	0.9072	0.9291	1978	0.9667	0.9689	0.9654
1968	0.6676	0.9183	0.9358	1979	0.9285	0.9649	0.9631
1969	0.6444	0.9272	0.9408	1980	0.6613	0.9546	0.9568
1970	0.5883	0.9355	0.9455	1981	0.7903	0.9510	0.9550
1971	0.6869	0.9454	0.9513	1982	0.8770	0.9450	0.9518
1972	0.8545	0.9557	0.9575	1983	0.8635	0.9347	0.9460
1973	0.8991	0.9615	0.9609	1984	0.9175	0.9265	0.9418
1974	0.9906	0.9695	0.9658	1985	0.8625	0.9131	0.9343
1975	0.9992	0.9746	0.9687	1986	0.4861	0.8904	0.9211
1976	0.9961	0.9735	0.9681	1987	0.3657	0.8750	0.9127
1977	0.9880	0.9722	0.9674	1988	0.3608	0.8607	0.9051

Berechnen wir für beide Hyperparameter die gesuchte Wahrscheinlichkeit $P(a_t < 0|y^{1988})$ aus den Momenten $\mathbf{x}_{t|1988,2}$ und $\mathbf{P}_{t|1988,22}$ der Glättungsdichte:

$$P(a_t < 0|y^{1988}) = \Phi \left(-\frac{\mathbf{x}_{t|1988,2}}{\sqrt{\mathbf{P}_{t|1988,22}}} \right).$$

$\Phi(\cdot)$ bezeichnet die Verteilungsfunktion der Standardnormalverteilung. Diese Wahrscheinlichkeiten sind in Tabelle 3 für jedes einzelne Jahr eingetragen. Die Ergebnisse zeigen deutlich, wie stark diese Wahrscheinlichkeit vom gewählten Hyperparameter abhängt.

Der Wahl des „richtigen“ Hyperparameters kommt damit große Bedeutung zu, wenn wir aus einem dynamischen stochastischen Modell Aussagen ableiten, die wesentlich von der Varianz der Filterdichten beeinflusst werden. Wir werden dieses Problem in Abschnitt 4.5.1 weiterbehandeln.

3.3.2 Datensatz 2 - Fallstudie 1

Analysieren wir nun den Datensatz der SO₂-Emissionen mit einem Steady-State-Modell für Gamma-verteilte Prozesse (Modell 2).

Approximativer Posterior-Mode-Filter. Zu jedem Zeitpunkt wird das nichtlineare Modell um den bedingten Erwartungswert von $x_t|y^{t-1} = \hat{x}_{t-1|t-1}^F$ linearisiert. Nach entsprechenden Umformungen ergibt sich folgender Zusammenhang zwischen den ersten Momenten der Filterdichte zum Zeitpunkt $t-1$ und t , der wegen der Fallunterscheidung $x_{t-1} \leq 1$ und $x_{t-1} \geq 1$ bei der Transformation $g^{-1}(x_{t-1})$ in den Bereichen $\hat{x}_{t-1|t-1}^F \geq 1$ und $\hat{x}_{t-1|t-1}^F \leq 1$ unterschiedliche Gestalt besitzt:

$$\begin{aligned} \hat{x}_{t|t}^F &= \begin{cases} (1 - K_t^{(1)}(\hat{x}_{t-1|t-1}^F))\hat{x}_{t-1|t-1}^F + K_t^{(1)}(\hat{x}_{t-1|t-1}^F)y_t, & \hat{x}_{t-1|t-1}^F \geq 1, \\ (1 - K_t^{(2)})\hat{x}_{t-1|t-1}^F + K_t^{(2)}\left(\frac{y_t}{e^{\hat{x}_{t-1|t-1}^F - 1}} + \hat{x}_{t-1|t-1}^F - 1\right), & \hat{x}_{t-1|t-1}^F \leq 1, \end{cases} \quad (18) \\ \hat{P}_{t|t}^F &= \begin{cases} \frac{(\hat{P}_{t-1|t-1}^F + Q)\frac{1}{\alpha}(\hat{x}_{t-1|t-1}^F)^2}{\hat{P}_{t-1|t-1}^F + Q + \frac{1}{\alpha}(\hat{x}_{t-1|t-1}^F)^2}, & \hat{x}_{t-1|t-1}^F \geq 1, \\ \frac{(\hat{P}_{t-1|t-1}^F + Q)\frac{1}{\alpha}}{\hat{P}_{t-1|t-1}^F + Q + \frac{1}{\alpha}}, & \hat{x}_{t-1|t-1}^F \leq 1, \end{cases} \\ K_t^{(1)}(\hat{x}_{t-1|t-1}^F) &= \frac{\hat{P}_{t-1|t-1}^F + Q}{\hat{P}_{t-1|t-1}^F + Q + \frac{1}{\alpha}(\hat{x}_{t-1|t-1}^F)^2}, \\ K_t^{(2)} &= \frac{\hat{P}_{t-1|t-1}^F + Q}{\hat{P}_{t-1|t-1}^F + Q + \frac{1}{\alpha}}. \end{aligned}$$

Die beiden Momente sind für $\hat{x}_{t-1|t-1}^F = 1$ identisch. In beiden Bereichen der Transformation ist der Zusammenhang zwischen den ersten Momenten nichtlinear. Für den Bereich $\hat{x}_{t-1|t-1}^F \geq 1$ ist dieser Filter mit dem Kalman-Filter eines Steady-State-Modells unter Annahme einer Normalverteilung für y_t mit zeitvarianter Beobachtungsvarianz $R_t = \frac{1}{\alpha}(\hat{x}_{t-1|t-1}^F)^2$ identisch. Die Varianz der Filterdichte hängt für diesen Bereich vom beobachteten Prozeß über $\hat{x}_{t-1|t-1}^F$ ab. Für den Bereich $\hat{x}_{t-1|t-1}^F \leq 1$ ist eine solche Interpretation nur indirekt möglich, wenn wir die Varianz der Filterdichte betrachten. Diese hat die selbe Gestalt wie bei einem Steady-State-Modell für einen normalverteilten Prozeß mit zeitinvarianter Varianz $R = \frac{1}{\alpha}$ und hängt nicht von den Beobachtungswerten ab.

Aus (18) sehen wir, wie eine nicht-informative a-priori-Dichte gewählt werden kann. Für

$$\hat{x}_{0|0} = 1, \quad P_{0|0} = d \quad (19)$$

erhalten wir für $d \rightarrow \infty$:

$$\hat{x}_{1|1} = y_1, \quad P_{0|0} = \frac{1}{\alpha}.$$

Abbildung 5: Datensatz 2 - Fallstudie 1

Naive Schwankungsintervalle für den Level $\mu_t|y^t$
(oben: Posterior-Mode-Filter, unten: Integrationsfilter)

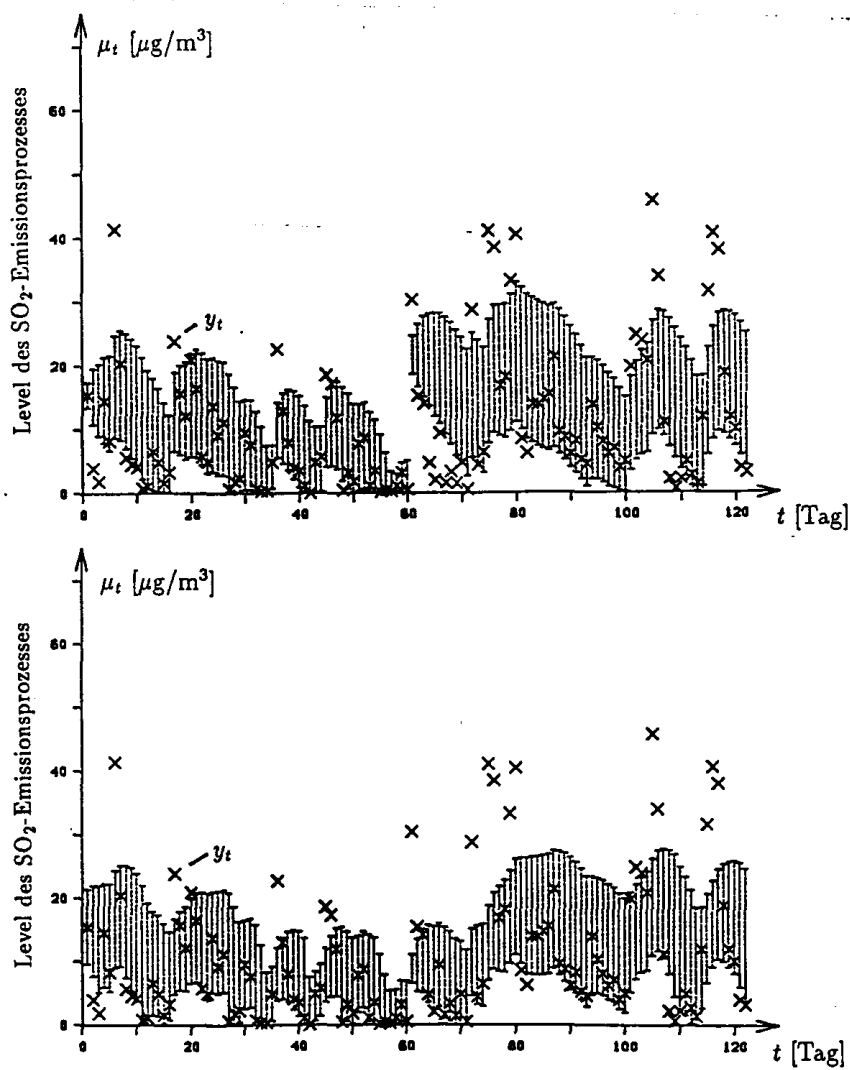
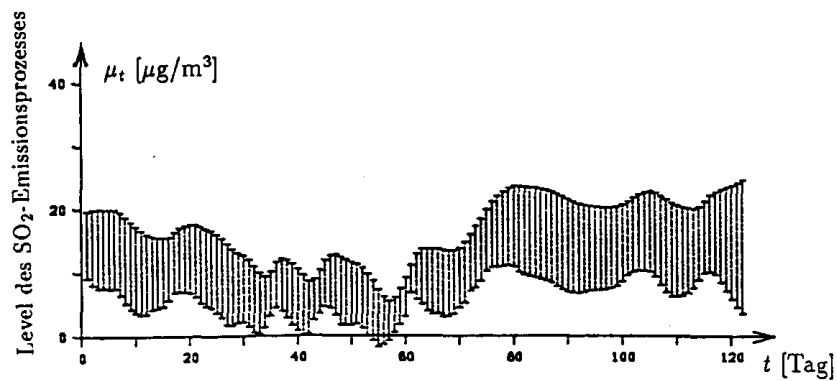


Abbildung 6: Datensatz 2 - Fallstudie 1

Naive Schwankungsintervalle für den Level $\mu_t|y^{122}$



Filtern für den Datensatz 2. Wir wählen für diesen Datensatz als Hyperparameter $\alpha = 1.11$ und $Q = 4$. Dieser Hyperparameter wurde aus den Daten mit Methoden geschätzt, die wir in Abschnitt 4 diskutieren und in Abschnitt 4.5.2 auf den Datensatz 2 anwenden werden. Der Parameter d der a-priori-Dichte (19) wurde gleich 10^4 gesetzt.

Wir vergleichen nun den approximativen Posterior-Mode-Filter mit einem Integrationsfilter mit $M_I = 10$ Stützstellen. Die beiden Charakteristika des Posterior-Mode-Filters und des Integrationsfilters sind für normalverteilte Filterdichten identisch. Unterschiede in den beiden Charakteristika bedeuten, daß die Gestalt der Filterdichte von der Normalverteilung abweicht. Zum Vergleich der beiden Filter wurden für jeden Zeitpunkt aus den beiden Charakteristika naive Schwankungsintervalle $S_{t|t}$,

$$S_{t|t} = [\hat{x}_{t|t} - 1.96 \cdot \sqrt{\hat{P}_{t|t}}, \hat{x}_{t|t} + 1.96 \cdot \sqrt{\hat{P}_{t|t}}],$$

ermittelt. Die Wahrscheinlichkeit, mit der $x_t|y^t$ in dieses Intervall fällt, beträgt 0.95, wenn die Filterdichte normalverteilt ist, ansonsten ist sie unbekannt. Durch Rücktransformation über $g^{-1}(\cdot)$ erhält man daraus ein naives Schwankungsintervall für den Erwartungswert $\mu_t|y^t$.

Abbildung 5 vergleicht diese Schwankungsintervalle für $\mu_t|y^t$ für beide Filter. Für die ersten 60 Beobachtungen sind die Intervalle nicht sehr verschieden. Der Posterior-Mode-Filter reagiert aber viel stärker auf die extreme Beobachtung bei $t = 61$ als der Integrationsfilter. Diese starke Reaktion des Posterior-Mode-Schätzers erklärt sich vermutlich aus der zu geringen Beobachtungsvarianz $0.909(\hat{x}_{60|60})^2$ der Normalverteilungsapproximation. Die Unterschiede zwischen den beiden Filtern nivellieren sich ab etwa $t = 85$.

Aussagen über den Verlauf des Erwartungswertes. Wir interessieren uns nun für Aussagen über den Verlauf des Erwartungswertes, um nach Strukturen in der Zeitreihe zu suchen. Unter Verwendung aller 122 Beobachtungswerte ermitteln wir nach dem Schema (11) für jeden Zeitpunkt $t = 121, 120, \dots, 1$ die Glättungsdichten $p(x_t|y^{122})$ ausgehend von den Filterdichten. Als Approximation der Momente der Filterdichte wählen wir die Charakteristika des Integrationsfilters.

Abbildung 6 zeigt naive Schwankungsintervalle, die für jedes t aus der Glättungsdichte bestimmt wurden. Als grobe Klassifizierung erhalten wir, daß der Verlauf des Erwartungswertes im September fallend war, im Oktober leichten zyklischen Schwankungen unterlag, im November wieder auf einen höheren Level anstieg, der für den Rest des Jahres annähernd konstant blieb.

3.4 Vorhersagedichten

3.4.1 Allgemeine Bemerkungen

Für dynamische stochastische Modelle ist die Vorhersagedichte $p(y_s|y^t)$ von künftigen Werten y_s des Prozesses über das Integral

$$p(y_s|y^t) = \int p(y_s|\mathbf{x}_s)p(\mathbf{x}_s|y^t)d\mathbf{x}_s \quad (20)$$

aus der Vorhersagedichte $p(\mathbf{x}_s|y^t)$ künftiger Werte des Zustandsvektors $\mathbf{x}_s|y^t$ bestimmbar. Die Vorsagedichte des Zustandsvektors $\mathbf{x}_s|y^t$ erhält man aus der dynamischen stochastischen Gleichung, die die Veränderung des Zustandsvektors beschreibt:

$$\begin{aligned} \mathbf{x}_s &= \mathbf{F}_s \mathbf{x}_{s-1} + \mathbf{w}_s, \quad \mathbf{w}_s \sim N(0, \mathbf{Q}_s) \Rightarrow \\ \mathbf{x}_s &= \mathbf{F}_{s|t} \mathbf{x}_t + \mathbf{w}_{s|t}, \quad \mathbf{w}_{s|t} \sim N(0, \mathbf{Q}_{s|t}), \quad s > t, \\ \mathbf{F}_{s|t} &= \mathbf{F}_s \cdot \mathbf{F}_{s-1|t}, \quad \mathbf{F}_{t|t} := \mathbf{I}, \\ \mathbf{Q}_{s|t} &= \mathbf{F}_{s-1} \mathbf{Q}_{s-1|t} \mathbf{F}_{s-1}^T + \mathbf{Q}_s, \quad \mathbf{Q}_{t|t} := 0. \end{aligned} \quad (21)$$

Für normalverteilte Prozesse sieht man aus dieser Darstellung, daß die Vorhersagedichte $p(\mathbf{x}_s|y^t)$ des Zustandsvektors \mathbf{x}_s die Dichte einer Normalverteilung ist, deren Momente sequentiell aus den Momenten der Filterdichte $p(\mathbf{x}_t|y^t)$ bestimmt werden können (siehe z.B. [32]):

$$\boxed{\mathbf{x}_t|y^t \sim N(\hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}) \Rightarrow \mathbf{x}_s|y^t \sim N(\hat{\mathbf{x}}_{s|t}, \mathbf{P}_{s|t})} \quad (22)$$

$$\begin{aligned} \hat{\mathbf{x}}_{s|t} &= \mathbf{F}_s \hat{\mathbf{x}}_{s-1|t}, \quad s = t+1, t+2, \dots, \\ \mathbf{P}_{s|t} &= \mathbf{F}_s \mathbf{P}_{s-1|t} \mathbf{F}_s^T + \mathbf{Q}_s. \end{aligned}$$

Für nicht-normalverteilte Prozesse ist die genaue Gestalt der Vorhersagedichte $p(\mathbf{x}_s|y^t)$ nicht bekannt, da von der Filterdichte nur bestimmte Charakteristika ermittelt wurden. Eine approximative, normalverteilte Vorhersagedichte für den Zustandsvektor entsteht, wenn die Filterdichte durch eine Normalverteilung mit diesen Charakteristika approximiert und mit dem normalverteilten Fehlerterm $\mathbf{w}_{s|t}$ in (21) überlagert wird. Die Momente dieser Vorhersagedichte werden nach demselben Schema (22) wie bei normalverteilten Prozessen berechnet.

Wenden wir uns nun der Vorhersagedichte zukünftiger Beobachtungen zu. Für normalverteilte Prozesse ist das Integral (20) analytisch lösbar, weshalb die Vorhersagedichte $p(y_s|y^t)$ zukünftiger Beobachtungen y_s die Dichte einer Normalverteilung ist, deren Momente mit den Momenten der Vorhersagedichte von $\mathbf{x}_s|y^t$ linear zusammenhängen (z.B. [32]):

$$\boxed{\mathbf{x}_s|y^t \sim N(\hat{\mathbf{x}}_{s|t}, \mathbf{P}_{s|t}) \Rightarrow y_s|y^t \sim N(\hat{y}_{s|t}, \mathbf{B}_{s|t})} \quad (23)$$

$$\begin{aligned} \hat{y}_{s|t} &= \mathbf{H}_s \hat{\mathbf{x}}_{s|t}, \\ \mathbf{B}_{s|t} &= \mathbf{H}_s \mathbf{P}_{s|t} \mathbf{H}_s^T + \mathbf{R}_s. \end{aligned}$$

Für nicht-normalverteilte Prozesse kann das Integral (20) auch dann nicht analytisch berechnet werden, wenn für die Filterdichte $p(\mathbf{x}_s|y^t)$ eine approximative Normalverteilung angenommen wird. Eine Reduktion der Dimension der Integration ist über die Transformation $\mu_s(\mathbf{x}_s) = g^{-1}(\mathbf{H}_s \mathbf{x}_s)$ möglich:

$$p(y_s|y^t) = \int p(y_s|\mu_s)p(\mu_s|y^t)d\mu_s. \quad (24)$$

μ_s besitzt eine Verteilung, von der nur die folgenden Momente bekannt sind:

$$E(g(\mu_s)|y^t) = \mathbf{H}_s \hat{\mathbf{x}}_{s|t} =: \lambda_{s|t}, \quad V(g(\mu_s)|y^t) = \mathbf{H}_s \mathbf{P}_{s|t} \mathbf{H}_s^T =: \Lambda_{s|t}. \quad (25)$$

Eine Approximationsmethode ([32]) besteht nun in der Wahl einer Verteilung $p(\mu_s|y^t)$, die zur Likelihoodfunktion $p(y_s|\mu_s)$ konjugiert ist. Das Integral (24) ist dann analytisch berechenbar. Die Parameter dieser approximativen Verteilung werden so gewählt, daß die Momentengleichungen in (25) erfüllt sind. Für die Identitätstransformation $g(\mu) = \mu$ sind diese Gleichungen einfach zu lösen, für die meisten anderen Transformationen muß ein nichtlineares Gleichungssystem in den Parametern gelöst werden.

Eine anderer Weg besteht in einer Übertragung der Ideen aus [25] und [26] auf die Vorhersage von nicht-normalverteilten Prozessen. Wir werden diese Methode in Abschnitt 3.4.2 beschreiben. Abschnitt 3.5 enthält Fallstudien zum Vorhersageproblem für die beiden Datensätze 1 und 2.

3.4.2 Vorhersage von nicht-normalverteilten Prozessen

Die Vorhersagedichte wird durch eine Summe von Dichten der Familie der Beobachtungsverteilung approximiert:

$$p(y_s|y^t) = \sum_{i=1}^{M_V} p(y_s|\mu_s^{(i)})w_{s|t}^{(i)}, \quad \sum_{i=1}^{M_V} w_{s|t}^{(i)} = 1. \quad (26)$$

Die einzelnen Dichten unterscheiden sich im Erwartungswert $\mu_s^{(i)}$. Die Erwartungswerte und die Gewichte ergeben sich aus einer eindimensionalen Gauß-Hermite-Integration von (24):

$$\mu_s^{(i)} = g^{-1} \left(m_s + \sqrt{2S_s} \tau^{(i)} \right), \quad w_{s|t}^{(i)} = \frac{\omega^{(i)} \cdot p_N(g(\mu_s^{(i)}); \lambda_{s|t}, \Lambda_{s|t})}{\sqrt{\pi} \cdot p_N(g(\mu_s^{(i)}); m_s, S_s)}.$$

$\lambda_{s|t}$ und $\Lambda_{s|t}$ sind die Momente (25). m_s und S_s werden so gewählt, daß das Gitter in einen Bereich fällt, über dem sich der Integrand konzentriert. Bei festem Argument y_s ist der Integrand eine unnormierte a-posteriori-Dichte. Wendet man, ähnlich wie in Abschnitt 3.2.1, eine lokale Linearisierung an, so erhält man approximative Momente des Integranden, die als Transformationsparameter m_s und S_s gewählt werden:

$$m_s = \lambda_{s|t} + K_s(\lambda_{s|t})(y_s - g^{-1}(\lambda_{s|t})),$$

$$K_s(\lambda_{s|t}) = \frac{(g^{-1})'(\lambda_{s|t})\Lambda_{s|t}}{\phi v_s(g^{-1}(\lambda_{s|t})) + \Lambda_{s|t}},$$

$$S_s = \frac{\phi v_s(g^{-1}(\lambda_{s|t}))\Lambda_{s|t}}{\phi v_s(g^{-1}(\lambda_{s|t})) + \Lambda_{s|t}}.$$

Diese Approximation ergibt für $M_V = 2$ die exakte Vorhersagedichte, wenn der Prozeß normalverteilt ist. Der Transformationsparameter m_s hängt vom Argument y_s ab, an dem die Vorhersagedichte berechnet werden soll, weshalb sowohl die Stützstellen als auch die Integrationsgewichte von y_s abhängen und für jeden Funktionswert y_s neu berechnet werden müssen. Dieser Nachteil kann vermieden werden, wenn als Transformationsparameter $m_s = \lambda_{s|t}$ gewählt wird. Die Exaktheit für normalverteilte Prozesse geht dann allerdings verloren.

3.5 Fallstudien zur Vorhersage

3.5.1 Datensatz 1 - Fallstudie 1

Versuchen wir, aufbauend auf den Ergebnissen aus Abschnitt 3.3.1, vorherzusagen, welchen Wert der mittlere Grundwasserspiegel in den Jahren nach 1988 annehmen wird. Die ersten Momente der Vorhersagedichte $p(y_{1988+l}|y^{1988})$ können auf einfache Weise aus den ersten Momenten der Filterdichte $p(\mathbf{x}_{1988}|y^{1988})$ ermittelt werden (siehe (22) und (23)):

$$\hat{y}_{1988+l|1988} = \hat{\mu}_{1988|1988} + l \cdot \hat{a}_{1988|1988}.$$

Für die zweiten Momente kann aus (22) und (23) folgende Darstellung der Varianz $B_{1988+l|1988}$ der Vorhersagedichte $p(y_{1988+l}|y^{1988})$ abgeleitet werden:

$$B_{1988+l|1988} = P_{1988|1988,11} + 2l \cdot P_{1988|1988,12} + l^2 \cdot P_{1988|1988,22} + \sum_{j=1}^l j^2 \cdot W + Q + R.$$

Aus der Vorhersagedichte lassen sich 95%-ige Prognoseintervalle $S_{s|t}^{0.95}$ ermitteln, deren

Tabelle 4: Datensatz 1 - Fallstudie 1

Prognoseintervalle für den mittleren Grundwasserspiegel

	1. Hyperparameter	2. Hyperparameter
$S_{1989 1988}^{0.95}$	[123.59, 124.60]	[123.59, 124.60]
$S_{1990 1988}^{0.95}$	[123.32, 124.93]	[123.33, 124.56]
$S_{1991 1988}^{0.95}$	[123.00, 125.32]	[123.09, 124.67]

Breite wesentlich von der Kovarianzmatrix der Filterdichte und vom gewählten Hyperparameter abhängt. Tabelle 4 enthält 95%-ige Prognoseintervalle für den mittleren Grundwasserspiegel in den Jahren 1989 – 1991 für beide Hyperparameter aus Abschnitt 3.3.1.

3.5.2 Datensatz 2 - Fallstudie 1

Betrachten wir für den Datensatz 2 folgendes Vorhersageproblem. Nehmen wir an, daß wir den SO_2 -Emissionsprozeß bis zum Tag t über die Tagesmittelwerte beobachtet haben. Wie groß ist dann die Wahrscheinlichkeit, daß der Tagesmittelwert der Schadstoffemissionen am nächsten Tag $t + 1$ einen gewissen Schwellwert S überschreitet? Die Antwort ergibt sich unmittelbar aus der Einschrittvorhersagedichte:

$$P(y_{t+1} > S|y^t) = \int_0^S p(y_{t+1}|y^t) dy_{t+1}.$$

Für einen festen Wert von y_{t+1} wird die Vorhersagedichte durch eine Summe von $M_V = 10$ Gamma-Dichten nach (26) approximiert, wobei $\lambda_{t+1|t}$ und $\Lambda_{t+1|t}$ aus den Momenten des Integrationsfilters aus Abschnitt 3.3.2 bestimmt werden:

$$\lambda_{t+1|t} = \hat{x}_{t|t}^I, \quad \Lambda_{t+1|t} = \hat{P}_{t|t}^I + Q.$$

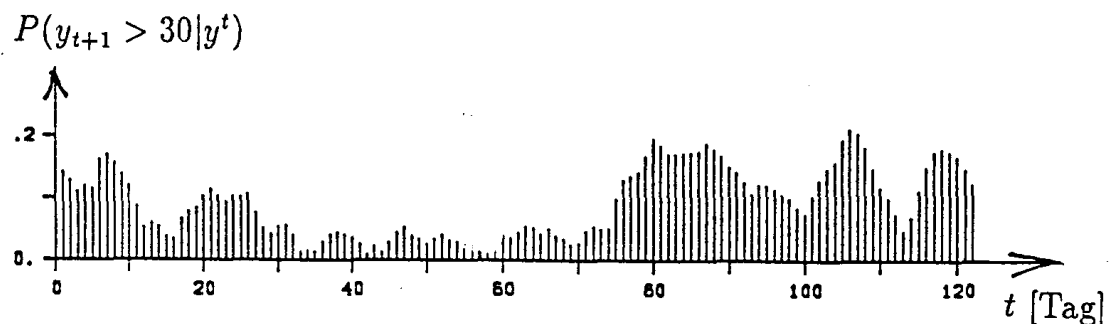
Zur Demonstration wurden diese Wahrscheinlichkeiten für den Schwellwert $S = 30 \mu\text{g}/\text{m}^3$ für jeden Tag berechnet und in Abbildung 7 graphisch dargestellt. Tabelle 5 enthält numerische Ergebnisse für ausgewählte Tage.

Tabelle 5: Datensatz 2- Prognosewahrscheinlichkeiten $P(y_{t+1} > 30|y^t)$
für die verschiedenen Fallstudien

t	6	16	27	44	50	63	75	80	100	106
Fallstudie 1	0.162	0.034	0.076	0.012	0.025	0.054	0.098	0.195	0.075	0.212
Fallstudie 2	0.159	0.037	0.076	0.016	0.027	0.052	0.094	0.204	0.076	0.213

Abbildung 7: Datensatz 2 - Fallstudie 1

Prognosewahrscheinlichkeiten $P(y_{t+1} > 30|y^t)$



4 Modelle mit unbekanntem Hyperparameter

4.1 Allgemeine Bemerkungen

Dynamische stochastische Modelle hängen im allgemeinen von einem Hyperparameter θ ab. Bei der Analyse der Grundwasserdaten in Abschnitt 3.3.1 wurde deutlich, daß der Hyperparameter erheblichen Einfluß auf die Gestalt der Filter-, Glättungs- und Vorhersagedichten, insbesondere auf deren Varianz, ausübt. Wir verwenden im weiteren die Bezeichnungen $p(\mathbf{x}_t|\theta, y^t)$, $p(\mathbf{x}_s|\theta, y^t)$ und $p(y_{t+l}|\theta, y^t)$ für diese Dichten, um die Abhängigkeit von θ sichtbar zu machen.

In der Praxis ist man mit dem Problem konfrontiert, wie der Hyperparameter zu wählen ist. Eine statistische Lösung dieses Problems besteht darin, θ ebenfalls aus der Zeitreihe zu schätzen (z.B. [20]). Allerdings entsteht dabei ein nichtlineares Schätzproblem, bei dem die Filter-, Glättungs- und Vorhersagedichten auch für normalverteilte Prozesse die Gestalt der Normalverteilung verlieren. Diese Dichten sind von folgender Gestalt:

$$p(\mathbf{x}_s|y^t) = \int p(\mathbf{x}_s|\theta, y^t)p(\theta|y^t)d\theta, \quad (27)$$

$$p(y_s|y^t) = \int p(y_s|\theta, y^t)p(\theta|y^t)d\theta. \quad (28)$$

$p(\theta|y^t)$ ist die a-posteriori-Dichte von θ bei gegebenen Daten y^t . Diese Dichte ist wegen des Bayes'schen Theorems proportional zum Produkt der Likelihoodfunktion von θ gegeben die Daten und einer a-priori-Dichte $p(\theta|y^0)$ (z.B. [20]):

$$p(\theta|y^t) \propto L(\theta|y^t)p(\theta|y^0). \quad (29)$$

Der Wert der Likelihoodfunktion $L(\theta|y^t)$ kann für dynamische stochastische Modelle sequentiell berechnet werden (siehe Abschnitt 4.2). Die a-posteriori-Dichte von θ ist im allgemeinen keine Dichte einer bekannten Verteilungsfunktion, weshalb man aus (29) nur den Funktionswert der nicht-normierten a-posteriori-Dichte für einen festen Wert θ berechnen kann.

Das Analyseproblem kann vereinfacht werden, wenn für θ ein Schätzwert $\hat{\theta}$, z.B. der Maximum-Likelihood-Schätzer ([13]) oder der Modus der a-posteriori-Dichte eingesetzt wird. Diese Maxima können im allgemeinen nur auf numerischem Wege, etwa durch direkte numerische Maximierung der Likelihoodfunktion (z.B. [13]) oder mit Hilfe des iterativen EM-Algorithmus ([7]) gefunden werden.

Zur Ermittlung der Filter-, Glättungs- und Vorhersagedichten wird der unbekannte Parameter durch den Schätzwert ersetzt, womit das Problem auf die Ermittlung dieser Dichten für voll spezifizierte Modelle reduziert wird (siehe Abschnitt 3):

$$p(\mathbf{x}_s|y^t) \approx p(\mathbf{x}_s|\hat{\theta}, y^t), \quad p(y_s|y^t) \approx p(y_s|\hat{\theta}, y^t).$$

Diese Vorgangsweise entspricht einer numerischen Integration von (27) und (28) mit einer Stützstelle in $\hat{\theta}$. Wenn die Dichte von $\theta|y^t$ nicht sehr konzentriert ist, wird diese Methode zu einer Unterschätzung der stochastischen Variation von $x_s|y^t$ und $y_s|y^t$ führen.

Eine Bayes'sche Lösung des Problems unbekannter Hyperparameter besteht in der Ermittlung der Filter-, Glättungs- und Vorhersagedichten unter Berücksichtigung der Unsicherheit in $\theta|y^t$. Dazu werden die exakten Dichten (27) und (28) durch Dichtesummen ersetzt (siehe Abschnitt 4.3). Diese Approximationsmethode wurde von [20] für normalverteilte Prozesse vorgeschlagen und von [14] Multi-Prozeß-Filter genannt. Die Erweiterung auf nicht-normalverteilte Prozesse ist offensichtlich.

Die Approximationseigenschaften des Multi-Prozeß-Filters hängen wesentlich von der Wahl einer guten Diskretisierung ab. Eine gute Diskretisierung liegt in einem Bereich, über dem sich die a-posteriori-Dichte $p(\theta|y^t)$ konzentriert. Eine solche Diskretisierung erhält man, wenn man die Hyperparameter $\theta^{(i)}$ aus der a-posteriori-Dichte $p(\theta|y^t)$ simuliert. Dazu ist es notwendig, zuerst die a-posteriori-Dichte $p(\theta|y^t)$ des Parameters θ zu ermitteln. Wir haben bereits erwähnt, daß diese Dichte keine geschlossene Form besitzt. Ähnlich wie der Suche des Maximum-Likelihood-Schätzers mittels des EM-Algorithmus kann diese Dichte durch ein iteratives Verfahren („Data-Augmentation“, [31]) approximiert werden (Abschnitt 4.4). Abschnitt 4.5 enthält Fallstudien zu den beiden Datensätzen 1 und 2.

4.2 Berechnung der Likelihoodfunktion

Der Wert der Likelihoodfunktion $L(\theta|y^t)$ kann für dynamische stochastische Modelle sequentiell berechnet werden, indem jede Beobachtung y_j , $j = 1, \dots, t$, in die Einschrittvorhersagedichte $p(y_j|\theta, y^{j-1})$ eingesetzt wird (z.B. [13]):

$$L(\theta|y^t) = \prod_{j=1}^t p(y_j|\theta, y^{j-1}) = L(\theta|y^{t-1})p(y_t|\theta, y^{t-1}).$$

Für normalverteilte Prozesse ist die Einschrittvorhersagedichte die Dichte einer Normalverteilung mit den Momenten $\hat{y}_{t|t-1}(\theta)$ und $B_{t|t-1}(\theta)$ – siehe (23) – und der Wert der Likelihoodfunktion einfach berechenbar.

Für nicht-normalverteilte Prozesse ist die Einschrittvorhersagedichte nicht analytisch berechenbar. Verwendet man den Filter auf Integrationsbasis, so erhält man *automatisch* den Beitrag $p(y_t|\theta, y^{t-1})$ der Beobachtung y_t zur Likelihoodfunktion, da dieser Wert mit der Integrationskonstanten $C_t(\theta)$ in (15) identisch ist, die durch $z_t(\theta)$ in (16) approximiert wird:

$$L(\theta|y^t) \approx \prod_{j=1}^t z_j(\theta). \quad (30)$$

4.3 Multi-Prozeß-Filter

Bei einem Multi-Prozeß-Filter ([20], [14]) werden die Dichten (27) und (28), die Mischungen aus unendlich vielen θ -bedingten Dichten sind, durch endliche Mischungen ersetzt:

$$p(\mathbf{x}_s|y^t) \approx \sum_{i=1}^{M_H} p(\mathbf{x}_s|\theta^{(i)}, y^t) w_t(\theta^{(i)}), \quad (31)$$

$$p(y_s|y^t) \approx \sum_{i=1}^{M_H} p(y_s|\theta^{(i)}, y^t) w_t(\theta^{(i)}). \quad (32)$$

Die Momente der bedingten Dichten $p(\mathbf{x}_s|\theta^{(i)}, y^t)$ und $p(y_s|\theta^{(i)}, y^t)$ ergeben sich für jedes $\theta^{(i)}$ mit den in Abschnitt 3 beschriebenen Methoden. Die Daten werden parallel mit M_H Modellen beschrieben, die sich im Hyperparameter unterscheiden. Aufbauend auf Beobachtungen bis zum Zeitpunkt t bewertet der Multi-Prozeß-Filter die einzelnen Modelle mit Gewichten $w_t(\theta^{(i)})$, die zum Wert der a-posteriori-Dichte an der Stelle $\theta^{(i)}$ proportional und normiert sind:

$$w_t(\theta^{(i)}) \propto p(\theta^{(i)}|y^t), \quad w_t(\theta^{(i)}) = \frac{p(\theta^{(i)}|y^t)}{\sum_{j=1}^{M_H} p(\theta^{(j)}|y^t)}.$$

Für normalverteilte Prozesse erhält man aus (31) und (32) eine Approximation der Filter-, Glättungs- und Vorhersagedichten durch eine Summe von Normalverteilungen, da die $\theta^{(i)}$ -bedingten Dichten Dichten einer Normalverteilung sind. Für nicht-normalverteilte Prozesse sind von $p(\mathbf{x}_s|\theta^{(i)}, y^t)$ für jedes $\theta^{(i)}$ nur bestimmte Charakteristika bekannt. $p(\mathbf{x}_s|y^t)$ kann ebenso wie bei normalverteilten Prozessen durch eine Summe von Normalverteilungen approximiert werden, wenn man die Momente der bedingten Dichten durch die Charakteristika des bedingten Filters ersetzt. Verwendet man zur Approximation der $\theta^{(i)}$ -bedingten Vorhersagedichte die Dichtesummenapproximation (26) aus Abschnitt 3.4.2 mit M_V Dichten, so führt die Approximation (32) der Vorhersagedichte $p(y_s|y^t)$ auf eine Summe von $M_H \cdot M_V$ Dichten der Familie der Beobachtungsverteilung.

4.4 Data Augmentation

Wendet man die Resulte von [31] auf dynamische stochastische Modelle an, so läßt sich die a-posteriori-Dichte $p(\theta|y^t)$ des Hyperparameters als Fixpunktlösung einer Integralgleichung darstellen:

$$p(\theta|y^t) = \int \int p(\theta|\mathbf{x}_0, \dots, \mathbf{x}_t, y^t) p(\mathbf{x}_0, \dots, \mathbf{x}_t|\theta', y^t) p(\theta'|y^t) d\mathbf{x}_0 \dots d\mathbf{x}_t d\theta'. \quad (33)$$

Die a-posteriori-Dichte wird iterativ aus (33) bestimmt: ausgehend von einer Approximation $g_{n-1}(\theta)$ verwendet man diese Gleichung, um die Approximation zu verbessern. Dieses

Verfahren konvergiert unter gewissen Regularitätsbedingungen gegen die a-posteriori-Dichte des Hyperparameters. In [11] wurde bewiesen, daß diese Regularitätsbedingungen für eine ganze Klasse von dynamischen stochastischen Modellen für normalverteilte Prozesse, zu der auch das Steady-State-Modell und das dynamische Trendmodell gehören, erfüllt sind. Die praktische Implementierung erfordert die Anwendung von Monte-Carlo-Methoden. Für eine detaillierte Beschreibung dieses Verfahrens muß auf [11] verwiesen werden.

Zur Approximation der Filter-, Glättungs- und Vorhersagedichten werden M_H Hyperparameter aus der a-posteriori-Dichte $p(\boldsymbol{\theta}|y^t)$ simuliert und ebenso wie beim Multi-Prozeß-Filter M_H Modelle mit verschiedenen Hyperparametern auf die Daten angewendet. Da es sich um eine Monte-Carlo-Integration von (27) und (28) handelt, sind die Gewichte anders als beim Multi-Prozeß-Filter zu wählen.

Die Glättungs- und Vorhersagedichten $p(\mathbf{x}_s|y^t)$ und $p(y_s|y^t)$ sind Mischungen mit den Gewichten

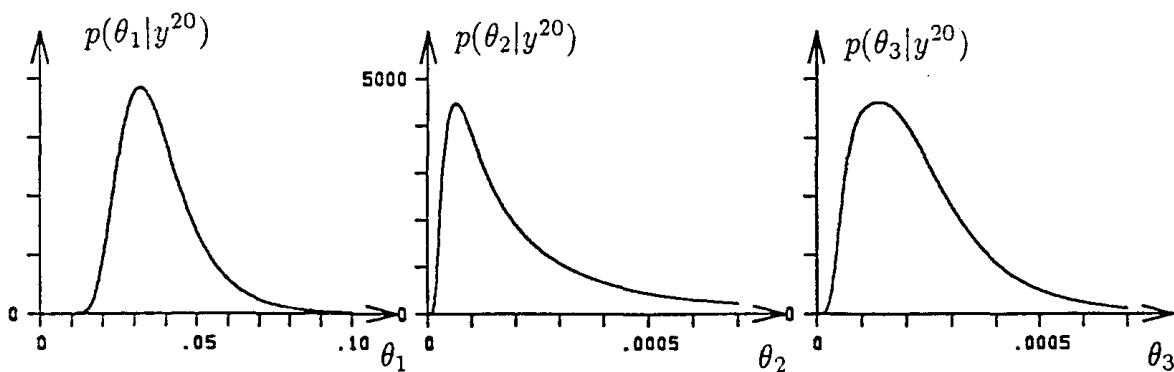
$$w_t(\boldsymbol{\theta}^{(i)}) = \frac{1}{M_H}. \quad (34)$$

Ist man an Dichten $p(\mathbf{x}_s|y^l)$ und $p(y_s|y^l)$ mit $l \neq t$ interessiert, so muß man die Gewichte folgendermaßen korrigieren:

$$w_l(\boldsymbol{\theta}^{(i)}) \propto \frac{p(\boldsymbol{\theta}^{(i)}|y^l)}{p(\boldsymbol{\theta}^{(i)}|y^t)}.$$

Abbildung 8: Datensatz 1 - Fallstudie 2

Marginale a-posteriori-Dichten der Komponenten des Hyperparameters



4.5 Fallstudien zum Problem unbekannter Hyperparameter

4.5.1 Datensatz 1 - Fallstudie 2

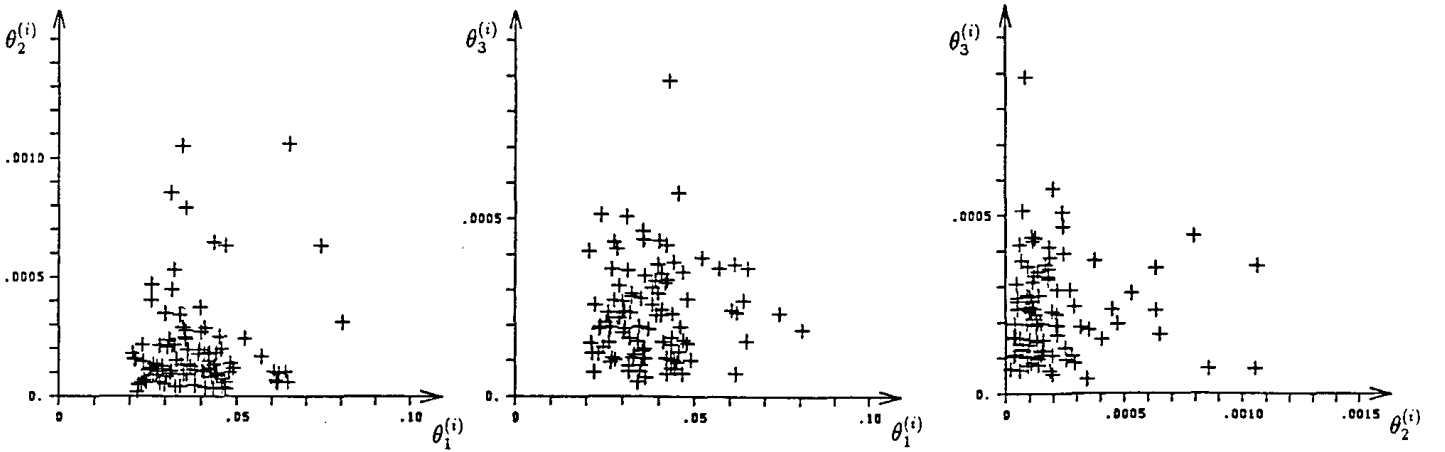
Wenden wir uns nun der Frage zu, wie der Hyperparameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (Q, W, R)$ zu wählen ist, wenn das Modell 1 auf den Datensatz 1 angewendet wird. Dieses Pro-

blem wurde in [11] ausführlich behandelt. Aufbauend auf einer wenig informativen a-priori-Dichte wurde in [11] die a-posteriori-Dichte des Hyperparameters mit dem Data-Augmentation-Algorithmus (Abschnitt 4.4) approximiert. Als Startapproximation $g_0(\theta)$ wurde eine Dichte gewählt, deren Erwartungswert mit dem ersten Hyperparameter aus Fallstudie 3.3.1 identisch ist. Nach 40 Iterationen ergab sich eine Approximation, deren Randdichten in Abbildung 8 graphisch dargestellt sind.

Aus dieser Approximation wurden 100 Hyperparameter simuliert (siehe Abbildung 9) und ein Multi-Prozeß-Filter mit Monte-Carlo-Gewichten (34) zur Bestimmung der Glättungsdichten $p(\mathbf{x}_t|y^{1988})$ angewendet. Der zweite Hyperparameter aus Abschnitt 3.3.1 ist einer dieser Gitterpunkte.

Abbildung 9: Datensatz 1 - Fallstudie 2

Monte-Carlo-Gitter für den Hyperparameter



Betrachten wir nun den Einfluß, den die Berücksichtigung der Unsicherheit in der Wahl der Hyperparameter auf Wahrscheinlichkeitsaussagen wie $P(a_t < 0|y^{1988})$ ausübt. Die Wahrscheinlichkeit $P(a_t < 0|y^{1988})$ ergibt sich wegen der Dichtesummenapproximation (31) mit den Monte-Carlo-Gewichten (34) als Mittelwert der Unterschreitungswahrscheinlichkeiten $P(a_t < 0|\theta^{(i)}, y^{1988})$ der $\theta^{(i)}$ -bedingten Glättungsdichten. Diese Wahrscheinlichkeit ist für jedes Jahr in Tabelle 3 den Ergebnissen aus Abschnitt 3.3.1 gegenübergestellt. Ein Vergleich mit dem zweiten Hyperparameter zeigt, daß sich diese Wahrscheinlichkeit nur um einige Prozent ändert.

4.5.2 Datensatz 2 - Fallstudie 2

Wie ist der Hyperparameter $\theta = (Q, \alpha)$ zu wählen, wenn der Datensatz 2 mit dem Modell 2 analysiert wird? Da der Data-Augmentation-Algorithmus bisher theoretisch nur für normalverteilte Prozesse untersucht wurde, arbeiten wir beim Gamma-verteilten Prozeß der SO_2 -Emissionen mit einem Multi-Prozeß-Filter, dessen Hyperparameter $(\theta)^{(i)}$ heuristisch ausgewählt wurden. Die Likelihoodfunktion ergibt sich für jeden Hyperparameter

aus den Normierungskonstanten eines Integrationsfilter mit $M_I = 10$ Stützstellen nach (30).

Tabelle 6: Datensatz 2 - Fallstudie 2

Multi-Prozeß-Filter mit grobem Gitter

$w_{122}(Q, \alpha)$	$Q = 0$	$Q = 1$	$Q = 5$	$Q = 10$
$\alpha = 1$	0.018	0.708	0.245	0.023
$\alpha = 2$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$
$\alpha = 5$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$

Tabelle 7: Datensatz 2 - Fallstudie 2

Multi-Prozeß-Filter mit feinem Gitter

$w_{122}(Q, \alpha)$	$Q=0.1$	$Q=0.5$	$Q=1$	$Q=2$	$Q=3$	$Q=4$	$Q=5$	$Q=6$	$Q=8$	$Q=10$
$\alpha = 1$	$< 10^{-3}$	0.005	0.020	0.049	0.069	0.072	0.055	0.032	0.011	0.004
$\alpha = 1.11$	$< 10^{-3}$	0.003	0.022	0.072	0.115	0.127	0.119	0.078	0.034	0.016
$\alpha = 1.25$	$< 10^{-3}$	$< 10^{-3}$	0.004	0.007	0.016	0.028	0.010	0.010	0.009	0.010

Die Auswahl der Hyperparameter erfolgte zunächst nach einem groben Gitter (siehe Tabelle 6). In dieser Tabelle sind auch die a-posteriori-Gewichte $w_{122}((\theta)^{(i)})$ der einzelnen Hyperparameter eingetragen, die auf gleichverteilten a-priori-Gewichten und allen 122 Beobachtungswerten beruhen. Die Modelle mit der schiefsten Beobachtungsverteilung ($\alpha = 1$) erhalten die größten Gewichte. Wir verfeinern nun das Gitter auf 30 Hyperparameter. Die a-posteriori-Gewichte dieses Gitters, die wieder auf gleichverteilten a-priori-Gewichten und allen 122 Beobachtungswerten beruhen, sind in Tabelle 7 zusammengefaßt. Aus dieser Tabelle kann man erkennen, daß wir für die Fallstudie in Abschnitt 3.3.2 den Hyperparameter mit dem größten a-posteriori-Gewicht gewählt haben. Dieser Parameter ist eine grobe Näherung des Maximum-Likelihood-Schätzers.

Wenden wir uns nun dem Einfluß zu, den die Berücksichtigung der Unsicherheit in der Wahl der Hyperparameter auf die Wahrscheinlichkeitsaussagen beim Prognoseproblem aus Abschnitt 3.5.2 ausübt. Die Vorhersagedichte ist eine Gewichtung der $\theta^{(i)}$ -bedingten Vorhersagedichten. Diese gesuchte Wahrscheinlichkeit ergibt sich daher durch Gewichtung der $\theta^{(i)}$ -bedingten Wahrscheinlichkeiten:

$$P(y_{t+1} > S|y^t) \approx \sum_{i=1}^{M_H} w_t(\theta^{(i)}) P(y_{t+1} > S|y^t, \theta^{(i)}).$$

Diese Wahrscheinlichkeiten werden in Tabelle 5 mit den Ergebnissen aus Fallstudie 3.5.2 für einzelne Zeitpunkte verglichen. Der Unterschied in den Werten beträgt nur einige Prozent.

5 Modelldiagnose

5.1 Allgemeine Bemerkungen

In diesem Abschnitt möchten wir kurz auf Methoden zur Modelldiagnose hinweisen. Betrachten wir zunächst vollspezifizierte Modelle für normalverteilte Prozesse. Zu jedem Zeitpunkt kann eine Einschrittvorhersage – charakterisiert durch die beiden ersten Momente $\hat{y}_{t|t-1}$ und $B_{t|t-1}$ der Vorhersagedichte $p(y_t|y^{t-1})$ (siehe (23)) – erstellt werden. Ein dynamisches stochastisches Modell für einen normalverteilten Prozeß wird überprüft, indem zu jedem Zeitpunkt die Vorhersagescores

$$e_t = \frac{y_t - \hat{y}_{t|t-1}}{\sqrt{B_{t|t-1}}} \quad (35)$$

berechnet werden (z.B. [21]). Die Vorhersagescores sind unabhängige Realisationen einer Standardnormalverteilung, wenn der Prozeß durch das gewählte Modell generiert wurde.

Für nicht-normalverteilte Prozesse sowie für Modelle mit unbekanntem Hyperparameter ist die Vorhersagedichte nicht normalverteilt. Vorhersagescores der Form (35) sind weder unabhängig noch normalverteilt. Eine Verallgemeinerung wurde von [24] vorgeschlagen, die darauf beruht, Scores der Unterschreitungswahrscheinlichkeiten (P-Scores) aus der Verteilungsfunktion der Einschrittvorhersagedichte $p(y_t|y^{t-1})$ abzuleiten:

$$u_t = \int_{-\infty}^{y_t} p(y|y^{t-1}) dy \quad (36)$$

Die P-Scores sind unabhängige Realisationen einer $[0,1]$ -Gleichverteilung, wenn der Prozeß durch das gewählte Modell generiert wurde ([24]). Für vollspezifizierte Modelle eines normalverteilten Prozesses hängen die P-Scores mit den Vorhersagescores (35) über $u_t = \Phi(e_t)$ zusammen. Für alle anderen Fälle ist die Berechnung der P-Scores aufwendiger. Für nicht-normalverteilte Prozesse verwendet wird die approximative Dichte (26) zur Berechnung der P-Scores. Für Modellen mit unbekanntem Hyperparameter wird die Dichtesummenapproximation (31) zur Approximation der P-Scores herangezogen.

Die Modelldiagnose besteht in der Analyse der P-Scores, etwa durch einen graphischen Vergleich ihrer empirischen Verteilungsfunktion mit der Verteilungsfunktion der Gleichverteilung. Eine statistische Modellprüfung ist möglich, wenn man testet, ob die empirischen P-Scores bestimmte unter der Annahme der Gültigkeit des Modells theoretisch zu erwartende Eigenschaften besitzen (siehe z.B. [28]).

5.2 Modelldiagnose mittels P-Scores für die Fallstudien dieser Arbeit

Betrachten wir zum Abschluß unserer Ausführungen die empirische Verteilungsfunktion der P-Scores für die vier Fallstudien dieser Arbeit (siehe Abbildung 10). Bei keiner dieser

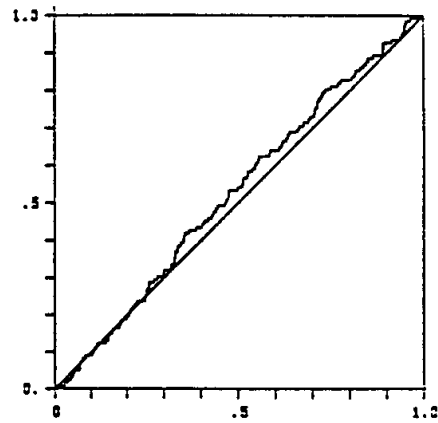
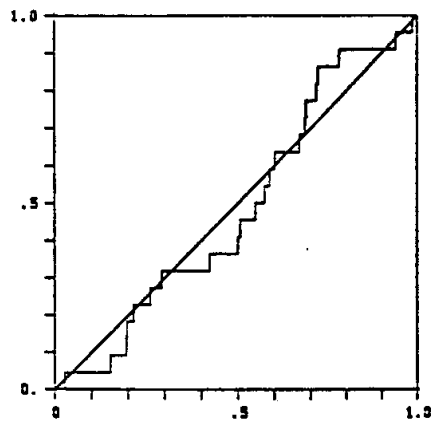
Abbildung 10: Empirische Verteilungsfunktion der P-Scores
aller Fallstudien dieser Arbeit

Datensatz 1

Datensatz 2

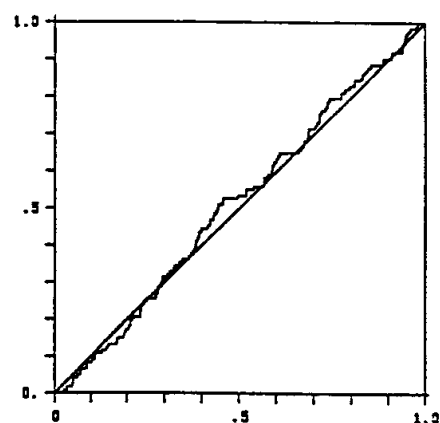
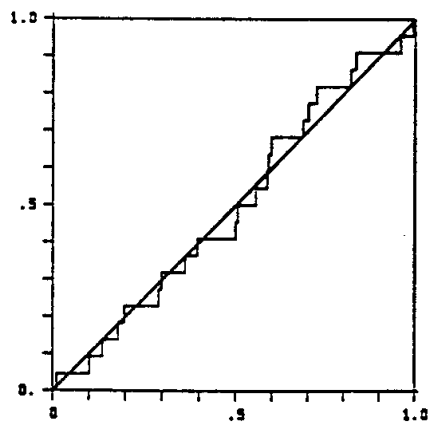
Fallstudie 1 (1. Hyperparameter)

Fallstudie 1 (Posterior Mode Filter)



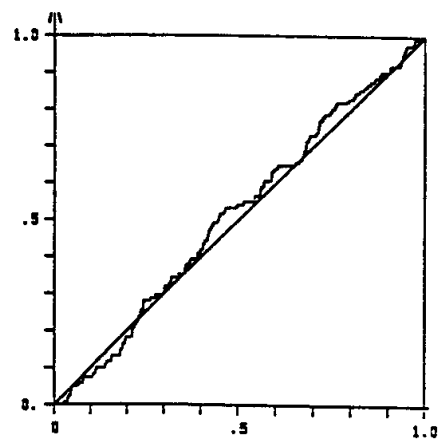
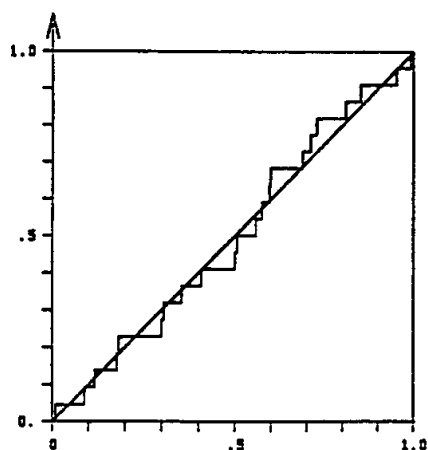
Fallstudie 1 (2. Hyperparameter)

Fallstudie 1 (Integrationsfilter)



Fallstudie 2

Fallstudie 2



Fallstudien ist aus diesen Diagrammen eine signifikante Abweichung zwischen der empirischen Verteilungsfunktion der P-Scores und der theoretisch zu erwartenden Gleichverteilung zu erkennen.

Bei Datensatz 1 ist allerdings offensichtlich, daß die Wahl von Hyperparametern, die aus den Daten geschätzt werden, die maximalen Abweichungen reduziert. Bei Datensatz 2 ist zu sehen, daß der Integrationsfilter Residuen produziert, deren empirische Verteilung näher an einer Gleichverteilung liegt, als die des Posterior-Mode-Filters.

Literaturverzeichnis

- [1] Abramowitz, M. u. Stegun, I.: *Handbook of Mathematical Functions*. National Bureau of Standards, New York, 1970.
- [2] Akaike, H.: Canonical Correlation Analysis of Time Series and the Use of an Information Criterion. In: R.K. Mehra u. D.G. Lainiotis (Hrsg.), *Advances and Case Studies in System Identification*, 27-96. Academic Press, New York, 1976.
- [3] Anderson, B.O.D. u. Moore, J.B.: *Optimal Filtering*. Englewood Cliffs, Prentice Hall, 1979.
- [4] Beck, M.B.: Water Quality Modeling – A Review of the Analysis of Uncertainty. *Water Resources Research*, 1983, **23**, 8, 1393-1442.
- [5] Beck, M.B. u. Van Straten, G. (Hrsg.): *Uncertainty and Forecasting of Water Quality*. Springer, Heidelberg, 1983.
- [6] Chiu, Ch. (Hrsg.): *Applications of Kalman-Filter to Hydrology, Hydraulics, and Water Resources*. Proceedings of AGU Chapman Conference. University of Pittsburgh, 1978.
- [7] Dempster, A.P., Laird, N.M. u. Rubin, D.B.: Maximum Likelihood from Incomplete Data Via the EM Algorithm. *J. R. Stat. Soc.*, 1976, **B**, **39**, 1-38.
- [8] Fahrmeir, L.: Extended Kalman-Filtering for Non-normal Longitudinal Data. In: A. Decarli, B.J. Francis, R. Gilchrist, G.U.H. Seeber (Hrsg.), *Statistical Modelling. Lecture Notes in Statistics*, **57**, 151-156. Springer, Berlin/ Heidelberg, 1989.
- [9] Fahrmeir, L. u. Kaufmann, H.: On Kalman-Filtering, Posterior Mode Estimation and Fisher-Scoring in Dynamic Exponential Family Regression. *Metrika*, 1991, **38**, 37 - 60.
- [10] Frühwirth-Schnatter, S.: Das Unbeobachtbare erfassen - Dynamische stochastische Modelle in den Umweltwissenschaften. In: R. Viertl (Hrsg.), *Beiträge zur Umweltstatistik*. Schriftenreihe der Technischen Universität Wien, wird voraussichtlich im Herbst 1991 erscheinen.

- [11] Frühwirth-Schnatter, S.: Approximating Posterior Densities for Dynamic Linear Models with Unknown Hyperparameters. Beitrag zum 6th International Workshop on Statistical Modelling, Utrecht 1991, Preprint.
- [12] Schilling, W. (Hrsg.): *Anwendungsmöglichkeiten des Kalman-Filter-Verfahrens in der Wasserwirtschaft*. VCH Verlagsgesellschaft, Weinheim/New York, 1987.
- [13] Harvey, A.: *Forecasting, Structural Time Series Models and the Kalman-Filter*. University Press, Cambridge, 1989.
- [14] Harrison, P.J. u. Stevens, C.F.: Bayesian Forecasting (with discussion). *J. R. Stat. Soc.*, 1976, **B**, **38**, 205-247.
- [15] Kalman, R.E.: A New Approach to Linear Filtering and Prediction Problems. *Trans. ASME, J. Basic Eng.*, 1960, **82**, 35-44.
- [16] Kitagawa, G.: Non-Gaussian State Space Modelling of Nonstationary Time Series (with comments). *JASA*, 1987, **82**, 1032-1063.
- [17] McCullagh, P. u. Nelder, J.A.: *Generalized Linear Models*. 2. Auflage. Chapman and Hall, London/New York, 1989.
- [18] Naylor, J.C. u. Smith, A.F.M.: Application of a Method for the Efficient Computation of Posterior Distributions. *Applied Statistics*, 1982, **31**, 214-225.
- [19] Crawford, M.M.: *Kalman Filters as an Enforcement Tool of the Air Quality Regulation*. Dissertation, University of Los Angeles, 1981.
- [20] Magill, D.T.: Optimal Adaptive Estimation of Sampled Stochastic Processes. *IEEE-TAC*, 1965, **10**, 434-439.
- [21] Mehra, R.K.: Aspects of Designing Kalman Filters. In: Ch. Chiu (Hrsg.), *Applications of Kalman-Filter to Hydrology, Hydraulics and Water Resources*, 89-114. Proceedings of AGU Chapman Conference. University of Pittsburgh, 1978.
- [22] Smith, A.F.M. u. West, M.: Monitoring Renal Transplants: an Application of the Multiprocess Kalman Filter. *Biometrics*, 1983, **39**, 897-878.
- [23] Smith, A.F.M., Skene, A.M., Shaw, J.E.H., Naylor, J.C., u. Dransfield, M.: The Implementation of the Bayesian Paradigma. *Communications in Statistics – Theory and Methods*, 1985, **14**, 1079-1102.
- [24] Smith, J.Q.: *Diagnostic Check of Non-standard Time Series Models*. Warwick University, Research Report **61**, Department of Statistics, 1985.
- [25] Schnatter, S.: *Dynamische Bayes'sche Modelle und ihre Anwendung zur hydrologischen Kurzfristvorhersage*. Unveröffentlichte Dissertation an der Technischen Universität Wien. Wien, 1988.

- [26] Schnatter, S.: Bayesian Forecasting of Time Series using Gaussian Sum Approximations. In: J.M. Bernardo, M.H. DeGroot, D.V. Lindley u. A.F.M. Smith (Hrsg.), *Bayesian Statistics 3*, 757-764. University Press, Oxford, 1988.
- [27] Schnatter, S.: Approximate Inference with a Dynamic Generalized Linear Trend Model. Preprint, zur Veröffentlichung in *Computational Statistics and Data Analysis* angenommen.
- [28] Schneider, W.: *Der Kalmanfilter als Instrument zur Diagnose und Schätzung variabler Parameter in ökonometrischen Modellen*. Physica, Heidelberg/Wien, 1986.
- [29] Schweppe, F.C.: Model Identification Problems. In: Ch. Chiu (Hrsg.), *Applications of Kalman-Filter to Hydrology, Hydraulics and Water Resources*, 115-133. Proceedings of AGU Chapman Conference. University of Pittsburgh, 1978.
- [30] Sharefkin, M.: Reflections of an Ignorant Bayesian. In: M.B. Beck u. G. van Straten (Hrsg.), *Uncertainty and Forecasting of Water Quality*, 373-379. Springer, Heidelberg, 1983.
- [31] Tanner, M. u. Wong, W.H.: The Calculation of Posterior Distributions by Data Augmentation. *JASA*, 1987, **83**, 398, 528-550.
- [32] West, M. u. Harrison, P.J.: *Bayesian Forecasting and Dynamic Models*. Springer, New York/Heidelberg/Berlin, 1989.
- [33] West, M., Harrison, P.J. u. Migon, H.S.: Dynamic Generalized Linear Models and Bayesian Forecasting. *JASA*, 1985, **80**, 389, 73-97.